Discovery of small molecules inhibitors of EphA4

Proefschrift

Ter verkrijging van

de graad van Doctor aan de Universiteit Leiden

op gezag van Rector Magnificus Prof. Mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 13 september 2012

klokke 16:15 uur

door

Carine Farenc

Geboren te Albi (Frankrijk) in 1981

PROMOTIECOMMISSIE

Promotor:	Prof. Dr. M. Ubbink
Co-Promotor:	Dr. G. Siegal
Overige leden:	Prof. Dr. J. Brouwer
	Prof. Dr. A. P. IJzerman
	Prof. Dr. R. Leurs
	Prof. Dr. T. Sixma

The work presented in this thesis has been financially supported by the Dutch Technology Foundation (Stichting Technische Wetenschappen, STW)

This thesis was printed by Wöhrmann Print Service (Zutphen, The Netherlands).

ISBN:978-94-6203-059-6

Cover design: Carine Farenc

“Je sers la science et c’est ma joie”

À ma famille

Contents

Abbreviations

Abl	Ableson kinase
ADME-Tox	absorption, distribution, metabolism, excretion, and toxicity
BG	benzylguanine
PEG	polyethyleneglycol
BTK	Bruton's tyrosine kinase
CSPs	chemical shift perturbations
ClogP	calculated logarithm of the octanol-water partition coefficient
DMSO	dimethyl sulfoxide
Eph	erythropoietin-producing hepatocellular
FBDD	fragment-based drug discovery
GPCR	G protein-coupled receptors
GPI	glycosylphosphatidylinositol
GST	glutathione S-transferase
HIV	human immunodeficiency virus
HSQC	heteronuclear single quantum coherence
HTS	high throughput screening
ILOEs	Interligand NOEs
INPHARMA	Interligand NOEs for PHARmacophore MApping
JMS	juxtamembrane segment
KD	kinase domain
LB	lysogeny broth
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser Effect
NTA	nitriolotriacetic acid
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCS	pseudo contact shift
PIN	percentage of inhibition
PRE	paramagnetic relaxation enhancement
RTK	receptor tyrosine kinase
RU	response units
SAR	structure-activity relationship
SPR	surface plasmon resonance
STD	saturation transfer difference
TEV	tobacco etch virus
TINS	target immobilized NMR screening
TMA	tetramethylammonium chloride
TPSA	total polar surface area
TROSY	transverse relaxation optimized spectroscopy
TSP	trimethylsilyl-2,2,3,3-tetradeuteropropionic acid

Chapter 1

General introduction

History of Drug Discovery

From the medicinal use of plants to modern drug discovery, the quest to discover healing drugs has always been associated with scientific research. In the 18th and earlier centuries, natural product extracts, particularly those derived from botanical species, proved the main source of folk medicine. During the Dark Ages the art of healing was intertwined with magic and religion; however Paracelsus in the 16th century was advocating the use of specific drugs to treat certain diseases.[1]

The beginning of the 19th century marks a critical junction owing to the advances made in the fields of chemistry and physiology. As pharmacologically active compounds were for the first time extracted from plants, patients began to receive a purified drug instead of a product of variable composition. In 1803 morphine was isolated from plants by Friederich Adam Sertürner.[2] Salicylic acid, the precursor of aspirin, was isolated in 1874 from willow bark. Various, more potent painkillers, such as codeine, were isolated from the opium poppy. Although synthesis of the first synthetic pharmaceutical drug, aspirin, occurred in the latter half of the nineteenth century, it was not until the early 1900s that the recognition of aspirin as a universal pain reliever was realized [3] and this discovery spawned the era of therapeutic agents.

The modern age of drug discovery began in 1937 with the discovery of the antibacterial action of sulfonamide. In 1943, Fleming discovered penicillin[4] and in 1947, streptomycin momentarily defeated tuberculosis.[5] The seeds for the concept of rational drug design were laid in the 1940s and 1950s by George Hitchings and Gertrude Elion in their work on DNA-based antimetabolites, which led to the discovery of modified purines

with anticancer activity.[6] However, the era of DNA and medicine was largely stimulated by the elucidation of the double-helical structure of DNA by Watson and Crick in 1953.[7] Exactly half a century – to the month – after this historical publication, the US Human Genome Project announced that the human genome sequence was "substantially complete".[8]

Around the mid-1970s began the era of recombinant DNA technology and molecular cloning: the first recombinant DNA molecules were generated and studied in 1972.[9, 10] Then in the 1980s and early 1990s, developments in molecular biology and virology had a major impact on the scientific understanding of the replication of the retrovirus, Human Immunodeficiency Virus (HIV). [11, 12] [11, 12] The polymerase chain reaction (PCR)[13] of the 1980s resulted in major advances in biotechnology that have had significant impact on drug discovery. In the 1990s, the concepts of combinatorial chemistry, molecular modeling and bioinformatics heralded a new age in drug discovery. The beginning years of the new millennium saw further advances in drug discovery based on state-of-the art chemistry, new advances in biology, enzyme-based molecular syntheses, proteomics and genomics, recombinant biomolecules, high-throughput screening and gene and cell therapy.[14]

Drug discovery process

The modern drug discovery process consists of sequential steps (Fig. 1.1). Once a target is identified and validated, a chemical lead is generated that is optimized into a

candidate. The lead candidate is then submitted to preclinical (animal) trials and eventually clinical trials.

Figure 1.1: The main steps in Drug Discovery.

This is not a straightforward process, in reality a large number of iterative cycles are involved in which a project will revolve until, finally, a promising molecular candidate is found that fits the profile needed to proceed to the next step.

- Target identification or target discovery attempts to identify principally proteins, whose modulation might inhibit or reverse the progression of a particular disease.[15]
- The validation step involves confirming the relevance of the target protein in a disease process. This step ideally requires both gain and loss of function studies using recent technological advances such as genomics, proteomics, small interfering RNA and mouse knockout models.[16]
- The hit discovery phase identifies compounds that interact with the target protein and modulate its activity. Principally, High throughput screening (HTS) is used to test large numbers of compounds for their ability to affect the activity of the target protein. Other methods such as Fragment-Based Drug Discovery and *in silico* screening are also employed to discover new chemical entities.

- During the optimization phase, small organic molecules are chemically modified and pharmacologically characterized in order to obtain compounds with suitable pharmacodynamics (efficacy, potency, selectivity *in vitro* and *in vivo*, physicochemical properties) and pharmacokinetics (ADME-Tox) properties to become a drug.

BOX 1: ADME-Tox

ADME-Tox is an acronym in pharmacokinetics and pharmacology for Absorption, Distribution, Metabolism, Excretion, and toxicity. It describes the disposition of a pharmaceutical compound within an organism. *In vitro* or *in silico* analysis of compounds during the lead optimization stage permits efficient prioritization of drug candidates. Most large pharmaceutical companies now routinely screen leads for *in vitro* toxicity and cytochrome p450 inhibition in order to minimize potential toxicity events later in drug development.

Hit discovery

The ability to identify active compounds that modulate a particular biomolecular pathway is a key step in drug discovery. Over the past decades, most pharmaceutical companies have used combinatorial chemistry and HTS resulting in the rapid identification of active molecules from large compound libraries.[17]

However, high attrition rates in the latter stages of drug development were observed, mainly due to selection of compounds with ADME-Tox characteristics (Box 1) that led to limited solubility, permeability and metabolic instability of compounds.[18] In addition, HTS has important limitations including large numbers of false positives[19] and false negatives[20] meaning the output of the screen requires detailed analysis and follow-up. Lipinski and colleagues examined a series of clinically tested drug molecules and

analyzed their properties. They introduced the concept of drug-likeliness with the Rule of Five[21] which constitutes a set of rules designed to predict whether or not a molecule is likely to be orally bioavailable *i.e.* has ideal ADME properties.

A further refinement of the concept of drug likeness was introduced : an analysis of drugs and their patented leads was made by Oprea and colleagues.[22] Based on the expected increase in molecular weight during the optimization process, it was suggested that the compounds used for screening should be lead-like[23] (Mw: 100-350 Da, and ClogP=1-3). A new approach for lead generation involving screening libraries of compounds significantly smaller and functionally simpler than drug molecules was developed: this approach is often referred to as "fragment-based" discovery.

Fragment Based Drug Discovery (FBDD) aims to create drug leads from small molecular building blocks. Since many of the building blocks were initially derived from marketed drugs, they have become known as drug fragments or simply fragments. Once appropriate fragments are identified, structure based methods are used to elaborate them towards potent and specific lead molecules. This technique was first introduced in 1981 by Williams Jencks who wrote that the affinities of whole molecules could be understood as a function of the affinities of separate parts.[24] This article spawned interest in the area of ligand-receptor interactions and provided a theoretical model for early work in which component fragments of biotin were found to bind weakly to streptavidin.[25] Further studies on HMG-CoA reductase inhibitors, showed that these inhibitors could be understood as a linkage of two fragments, each binding to distinct sites on the enzyme.[26] It was not until 1996 however, that the first practical demonstration of an FBDD method, called SAR by NMR, (structure-activity relationship by nuclear magnetic resonance) was

published by researchers at Abbott Laboratories.[27] In these studies, NMR was used to detect the weak interactions of the fragments with the target and determine the binding site.

The definition of a fragment varies, but usually refers to molecules weighing less than 300 Da, as proposed originally by Congreve and colleagues[28] in their "rule of three" or Ro3. The rule of three is defined as a compound with molecular weight < 300 Da with fewer than 15–20 heavy atoms, less than 3 hydrogen bond donors and 3 hydrogen bond acceptors, number of rotatable bonds < 3 and TPSA (total polar surface area) < 60 $Å^2$) and high aqueous solubility (ClogP ≤ 3).[29] The main advantage of FBDD is represented by the low molecular weight of the fragments of ligands used in screening for binding to a target protein. Due to the simplicity of the fragments it is possible to test a reasonable area of chemical space with small numbers of compounds. The number of potential molecules with up to 30 heavy atoms is estimated at more than 10^{60}, [30] whereas the number of potential fragments with up to 12 heavy atoms has been estimated at 10^7. [31] Thus, FBDD can screen a much greater proportion of the chemical space than HTS, hence increasing the chances to find new molecular entities.

In the past decade FBDD has become a powerful tool for discovering new leads. In 2007, 13 different institutions reported the development of more than 50 potent (IC50 <100 nM) inhibitors against diverse protein targets starting form weakly binding fragments.[32] To date, at least 18 drugs derived from FBDD have entered the clinic and 3 of them targeted protein kinases.[33] AT7519, an inhibitor of cyclin-dependent kinase 2 (CDK2)[34] and AT9283 targeting Aurora A kinase[35] were developed at Astex

Therapeutics by means of X-ray crystallography which was used to advance initial fragment hits to an experimental drug. The development of PLX4032, a selective inhibitor of the kinase B-Raf,[36, 37] began with a crystallographic screening of fragments and then selectivity and potency were improved during the course of optimization. It has been shown that the success of FBDD is significantly improved when 3D structure information is available.[32] Thus far, X-ray crystallography has been the primary structural tool driving FBDD.

As FBDD screens small molecules, they are correspondingly weaker than most hits from HTS,[38] with typical binding affinities in the range of 1 μM - 10 mM K_D. Accordingly, biophysical screening methods, including NMR and X-ray crystallography are particularly suitable, given the range of binding affinities they can detect. Recently, an NMR ligand based technique called Target Immobilized NMR Screening (TINS) was developed in our laboratory.[39] TINS as its name indicates, screens a collection of fragments for binding to an immobilized target using NMR. The method involves immobilization of a target protein and a reference protein on a commercially available resin. A flow-injection, dual-cell sample holder[40] can be placed in the magnet to enable tubeless, stop-flow screening of fragments by automated injection of mixtures containing up to 8 compounds into the immobilized protein samples.

Receptor tyrosine kinases and Cancer

Protein kinases constitute one of the largest and most functionally diverse gene families. They control many different cellular functions by phosphorylating proteins

involved in gene expression, metabolic pathways, cell growth and differentiation. Kinases are particularly prominent in signal transduction and coordination of complex functions such as the cell cycle. A number of diseases and disease symptoms result from the dysregulation of kinases or from mutations in kinases.[41]

More than 500 different protein kinases are found in the human genome including several subfamilies differentiated by the type of amino acid residue phosphorylated (Tyr, Thr or Ser). The family can also be divided into receptor and non-receptor kinases. Receptor tyrosine kinases (RTK's) are cell surface, transmembrane proteins responsible for the transduction of regulatory signals from the extra- to the intracellular milieu. RTKs are expressed in several cell types and are critical components in signal transduction pathways involved in cellular proliferation, differentiation, migration and metabolism.[42] Kinases such as the platelet-derived growth factor (PDGF) and epidermal growth factor (EGF) receptors are commonly activated in cancer cells and are known to contribute to tumorigenesis.[43] In most cases, gene amplification, translocation, mutation or overexpression are responsible for the acquired transforming potential of oncogenic RTKs.[44]

Anticancer approaches that function through the prevention or interception of dysregulated RTK signaling include the development of selective compounds that target the extracellular ligand binding domain or the intracellular tyrosine kinase domain. From the approval of the first kinase inhibitor, imatinib,[45] for the treatment of chronic myelogenous leukemia (CML) in 2001 to the approval of vandetanib[46] in early 2011, protein kinases have developed into one of the most important target classes in oncology drug discovery. However, studies revealed that most approved kinase inhibitors have

limited selectivity and target multiple kinases.[47] In addition, half of the inhibitors that have entered clinical trials target kinases that are already the target of approved drugs.[48] Thus, there is significant demand for breakthrough therapies directed towards previously untargeted kinases such as EphA4.

EphA4

The erythropoietin-producing hepatocellular (Eph) family of receptor tyrosine kinases are key regulators of diverse cellular functions including cell proliferation, differentiation and migration.[49] In addition to the roles that kinases are traditionally associated with, Eph receptors are also key factors in processes such as axon growth and angiogenesis.[50, 51] To date 14 different human Eph receptors and 8 different ligands, known as ephrins, have been identified.[52] The ephrins are divided into two subclasses A and B, based on their glycosylphosphatidylinositol (GPI)-anchored or transmembrane structure, respectively. The receptors are also classified into 2 families, EphA and EphB, depending on the binding affinity for either the A or B ephrins respectively.[53] The most well known exception to this general rule is EphA4, which binds both A- and B-class ligands.[54-56]

In addition to its normal role in cellular regulation, over expression of EphA4 has been observed in a variety of malignant carcinomas including gastric cancer[57] and cutaneous lymphomas[58] among others. Furthermore, the over-expression of EphA4 seems to be critical for tumor growth in a significant percentage of prostate cancers.[59] In addition to its potential role in oncology, data obtained from EphA4-knockout mice

suggest that blocking EphA4 function might be beneficial for the treatment of spinal cord injuries.[60, 61] The ability to modulate the activity of this family of receptors could prove useful and therefore they represent interesting therapeutic targets.

Research Project

The aim of the research project that is the subject of this thesis was to develop high affinity inhibitors of the tyrosine kinase EphA4. In light of the success of FBDD, particularly for developing kinase inhibitors, we sought to implement the paradigm in a multi-laboratory, collaborative effort to develop EphA4 inhibitors with optimized, drug-like characteristics. To reach this goal, a consortium consisting of three complementary academic groups and three Dutch companies was created. The principle investigators on the project were: Dr. Tensen from the department of dermatology at the Leiden University Medical Center (LUMC), Dr. de Esch from the Division of Medicinal Chemistry at the Vrije Universiteit (VU) and Dr. Siegal from the Leiden Institute of Chemistry (LIC). The companies involved in the project were: Pyxis Discovery, ZoBio and PamGene. (Fig. 2.2)

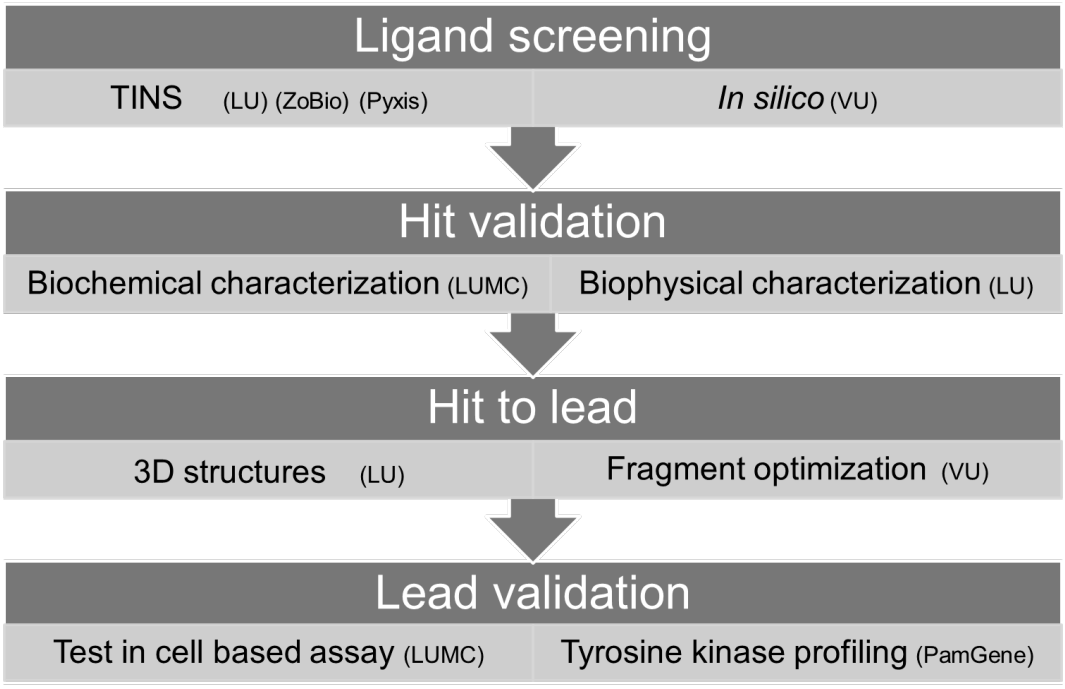


Figure 2.2: Overview of the research project organization. Main stages of the project are shown with the participants. The academic groups were: LU, Leiden University with Dr. Siegal, VU, Vrije University with Dr. de Esch and LUMC, Leiden University Medical Center with Dr. Tensen. The companies were ZoBio, Pyxis and PamGene.

The overall structure of the project is outlined in Figure 2.2. The ligand discovery was based on two complementary approaches: an NMR-based ligand screen and a computational screen. The NMR based ligand screening was performed at Leiden University in collaboration with ZoBio using TINS.[39] As input for ligand screening, a diversity library highly optimized for TINS was previously developed by Pyxis Discovery in collaboration with ZoBio.[62] *In silico* screening of the EphA4 kinase domain and design and synthesis of new ligands was performed by the Ph.D. candidate Oscar van Linden in the group of Dr. de Esch. Compounds were tested for biological function in the

group of Dr. Tensen and in parallel, biophysical characterization of the hits was performed at Leiden University. Structural studies of EphA4 were carried out with the support of the Nederlands Kanker Instituut (NKI-AVL) in Amsterdam.

Overview of the thesis

This thesis is divided into chapters that cover the discovery of small molecule inhibitors of EphA4. Chapter 2 presents an overview of biomolecular NMR in drug discovery. Chapter 3 describes the initial steps in ligand discovery and characterization including the screening of a fragment library for compounds that bind to the kinase domain of EphA4. Chapter 4 describes the crystallization and structure elucidation of the EphA4 kinase domain in the native state. In addition, the co-crystallization of this domain with a well known kinase inhibitor, Dasatinib, is presented. The *in silico* ligand screen performed by the group of Dr. de Esch in Amsterdam yielded an interesting and potent inhibitor of the EphA4 kinase domain. The discovery of this compound and the biochemical characterization as well as the crystal structure is described in Chapter 5. Biophysical characterization of the fragments discovered with TINS is presented in Chapter 6. Finally, in the concluding remarks the results from these chapters are put into a broader perspective and are discussed in relation to each other.

Chapter 2

Advances in Biomolecular NMR in Drug Discovery

Over the past decades nuclear magnetic resonance spectroscopy has been evolved into a mature technique in support of structure based drug design. A significant number of industrial and academic laboratories employ NMR to provide information on the nature of molecular interactions. In this Chapter, the major applications of NMR in different stages of drug discovery process are presented.

Farenc C. and Siegal G. *Recent Advances in Biomolecular NMR for Drug Discovery*. Chapter from "Recent Advances in Biomolecular NMR" in the RSC Biomolecular Sciences Series. (2012). http://www.rsc.org/shop/books/2012/9781849731201.asp

The use of NMR in industrial drug discovery has come and gone in waves over the years. In the 1990s significant investments in high field instruments and personnel were made on the basis of NMR as a tool to rival X-ray crystallography in the elucidation of 3D structures of protein-small molecule complexes. Unfortunately the field did not deliver within the timescales and throughput required to support commercial efforts and many NMR departments were reorganized. Subsequently the group at Abbott Laboratories lead by Stephen Fesik introduced the concept of using NMR to screen through collections of so-called drug fragments for binding to a protein target. Although the underpinnings of Fragment Based Drug Discovery (FBDD) can be traced back to academia,[63] the clear demonstration of the usefulness of NMR for commercial ends lead to substantial new interest. In the intervening 15 years NMR has become a mainstay of the fragment screening, validation and elaboration process.

The primary reason that NMR is so useful for characterizing fragments is the exquisite sensitivity of the technique to intermolecular interactions. This sensitivity is manifested in two different phenomena: the chemical shift and relaxation. The chemical shift with its intimate relationship to the immediate chemical environment of a nucleus is a ubiquitous probe that is readily and simply available. In practice the chemical shift is most often used in conjunction with NMR spectroscopy of the target (protein or target observed NMR), where it can not only differentiate specific, reversible binding from artifactual interactions, but, given the availability of the sequential assignment, can define an approximate binding site. Protein observed methods typically require isotopic labeling placing certain limitations on the applicability of the method. The very large difference in relaxation behavior between a small molecule and a protein is typically exploited to

characterize binding by observing changes in the NMR spectrum of the small molecule (ligand observed NMR). Ligand observed methods require substantially less protein, which need not be labeled, and may be more sensitive than protein observed methods, but may be more prone to artifacts.

Due to the demonstrated utility of NMR for drug discovery, there continues to be developments in the field that are both evolutionary and revolutionary. In this review we highlight selected recent accomplishments in the use of NMR for finding and characterizing small molecule ligands primarily during the very early stages of pre-clinical drug discovery. In addition, exciting developments in detecting molecules binding to targets inside living cells will also be presented. In this chapter, the emphasis is made on the newer technical development and functional insights obtained in recent years.

Table of contents

1. NMR for ligand discovery

At present ligand discovery and characterization comprises the major use of NMR in the drug discovery process. Despite the fact that NMR has been used for more than 15 years in this role, interesting new methods continue to be developed and exciting results continue to flow. Interestingly, a recent poll on Practical Fragments (http://practicalfragments.blogspot.com/) suggests that including both protein and ligand observed techniques, NMR is the most frequently used method of fragment screening.

1.1. Protein observed NMR

Protein observed screening is the only NMR method which provides information on the ligand binding site as well as the mode of binding to the protein. Despite the fact that it was introduced 15 years ago,[27] protein observed NMR remains the gold standard for identifying weak, yet specific binding of fragments to targets and differentiating these from non-specific or artifactual interactions. The very nature of the technique introduces some limitations such as the requirement for isotope labeling and the restriction to small/medium sized proteins. However, advances such as the introduction of TROSY,[64] particularly in conjunction with selective labeling schemes that introduce NMR visible isotopes only at e.g. methyl groups,[65] have allowed complete backbone assignment and ligand screening of considerably larger proteins.

Among target-based NMR techniques, the heteronuclear single quantum coherence (HSQC) spectrum of a ^{15}N or ^{13}C isotope labeled protein is by far the most commonly

used to detect ligand binding. This technique monitors chemical shift perturbations (CSPs) in the [^{15}N,^{1}H]-HSQC spectrum upon ligand binding. An application of this method to BACE-1, a protease that is a target for Alzheimer's disease, was recently published.[66] Since the location of BACE is within the brain, small molecule inhibitors must first traverse the blood brain barrier. This extra requirement places further restrictions on BACE inhibitors than other proteases, a class of targets already considered challenging. The group of Daniel Wyss developed an efficient scheme to produce labeled protein from inclusion bodies and used the protein to screen about 10,000 fragments before the sequential assignment was available. By first titrating known peptide analogs, the relevant CSPs were determined and hits from the screen were selected by their effect on these resonances. When the sequential assignment became available (*via* triple labeling and through bond methods),[67] binding to the active site was confirmed for 9 independent classes of compounds.[68] This latter publication also contains many suggestions for successful elaboration of initial hits discovered by NMR. The most promising fragment exhibited instability and therefore the HSQC experiment was used to screen for stable isosteres that also bound the active site aspartates of BACE-1. Although this search was not ultimately fruitful, structure activity relationship (SAR) information gathered led to the conceptualization of a new core which was also characterized by NMR. The ability of the HSQC experiment to quickly discern different binding modes of the BACE inhibitors proved to be critical for the project. The approach has successfully yielded a compound that is now in clinical trials.

In an interesting twist to HSQC screening, Holak and coworkers[69] developed the antagonist induced dissociation assay (AIDA) screen for inhibitors of protein-protein

interactions. This method monitors changes that occur when a large protein (more than 30 kDa) binds to a smaller protein (less than 20 kDa). When the complex is formed, the resonances broaden due to increased transverse relaxation. In its initial incarnation the smaller protein (the N-terminal p53 binding domain of MDM2) was isotopically labeled and mixed with the larger protein (the N-terminal 350 amino acid residues of p53) resulting in the disappearance of most of the MDM2 peaks. Addition of molecules that disrupt the protein-protein interaction restore the spectrum of MDM2. In contrast, if compounds simply bind one partner the spectrum is unchanged. Variants of this technique have been recently implemented: A 1D proton version of AIDA[70] and SEI-AIDA.[71] The 1D proton NMR version of AIDA monitors ligand binding though the effect on Trp proton resonances, alleviating the requirement for labeled protein. The SEI- AIDA (SEI for Selective Excitation Inversion) combines the 1D proton technique with a selective excitation of protein resonances which enables shorter relaxation delays.

Finally, protein observed NMR is commonly used to validate hits selected *via* a pre-screen. A nice illustration is provided by the virtual screening of the disheveled PDZ domain.[72, 73] Titration of 15 hits from the screen into ^{15}N-HSQC labeled PDZ showed that all hits bound the peptide binding site with the most potent in the low μM range. One of the hits blocked Wnt signaling in a cellular assay.

1.2. Ligand observed NMR

Ligand observed NMR is most commonly used amongst all screening techniques as it does not require any isotopic labeling and few limitations are imposed on the target.

Ligand based NMR techniques are mostly based on the difference in size between the ligand and the protein. The size difference manifests itself in numerous observables such as enhanced relaxation or a change in the diffusion coefficient. In addition, complex formation may be observed by transfer of magnetization from protein to fragment or by changes in the chemical shift or lineshape of ^{19}F spins in a ligand.

1.2.1. Relaxation methods

In NMR, relaxation is the process which restores equilibrium magnetization. Two different types of relaxations can be distinguished and can be characterized by their rate: longitudinal relaxation and transverse relaxation. Longitudinal relaxation is a complex function of molecular weight while transverse relaxation increases with molecular weight. Ligand observed NMR techniques using relaxation methods are based upon the fact that a ligand bound to the protein adopts the relaxation properties of the complex, i.e. transverse relaxation will be greatly enhanced.

As just described, larger, more slowly tumbling molecules relax much faster than small molecules. By extension, immobilization of the target on a solid support will further enhance transverse relaxation such that the difference with a small molecule in solution is at least 2 orders of magnitude. We have constructed an NMR fragment screening apparatus based on this principle which we call TINS for Target Immobilized NMR screening.[39] TINS uses differences in the spectrum of the small molecule in the presence of the target and a reference protein to detect binding. In a first, TINS was used to screen a fragment collection for specific binding to a membrane protein, DsbB that

forms part of a disulfide catalytic cascade in gram negative bacteria.[74] Membrane proteins are particularly challenging due to the very high level of false positives that arise from non-specific binding to the hydrophobic solubilization media. In this case the reference sample is used to cancel out the non-specific component of binding. More recently, we have also succeeded in screening a GPCR using TINS.[75]

NMR relaxation can also be enhanced by use of paramagnetism, i.e. the interaction between unpaired electrons and the nuclear spin. Two types of interaction can be defined depending on the nature of the paramagnetic center. Isotropic centers give rise to a purely distant dependent increase in the transverse relaxation rates of nuclear spins (paramagnetic relaxation enhancement or PRE) where anisotropic centers tend to cause a shift in the resonance frequency of spins (Pseudo contact shift or PCS). If a paramagnetic center (a spin label) is attached to a protein at a defined site, it is possible to measure distance dependant and/or angle dependant affects on the NMR spectrum of a ligand bound in the vicinity of the spin label. The first ligand screening technique to take advantage of a spin label was called SLAPSTIC (Spin Labels Attached to Protein Side chain as a Tool to identify Interacting Compounds).[76] The use of a spin label for screening not only enhances the sensitivity of the method, it enhances the specificity and is selective for a small binding site in proximity to the spin label. This latter property was put to use to find ligands specific for the hydrophobic binding pocket of the HIV-1 fusion protein gp41.[77] Here the spin label was attached to the terminus of an engineered α-helical peptide derived from the C-terminus of gp41. The peptide was truncated such that the hydrophobic binding site on the target remained open for small molecule binding. By measuring both PRE and PCS affects, the authors could determine structural constraints

to include in docking procedures leading to a model of the ternary complex. Similarly, a lanthanide binding peptide was fused to the SH2 domain of Grb2 where proof of concept data demonstrated the feasibility of the system for screening for weakly binding, small molecules.[78] The authors then used PCS to elucidate the structure of a complex of a phosphotyrosine peptide in fast exchange with the SH2 domain and a non-peptidic inhibitor in slow exchange. In both cases the correspondence between the structures determined using paramagnetic affects and one based either on X-ray crystallography or NOEs was reasonable. The power of the system lies in the specificity of the screen and the ability to rapidly get structural information from the same system.

The Pellecchia group has recently published[79] an interesting twist on paramagnetic labeling that is reminiscent of the earlier work from Novartis.[76] The group aimed to develop inhibitors of the protein tyrosine phosphatase of *Yersina pestis.* A spin labeled phosphotyrosine mimetic was developed and a library of about 500 low molecular weight compounds was screened for proximal binding. 8 positives from the screen were then linked to the phosphotyrosine mimetic to yield a series of novel bidentate compounds with low or even sub-μM IC_{50}s against the YopH phosphatase and activity in a cell-based assay.

1.2.2. Magnetization transfer methods

Method based on the transfer of magnetization from or *via* the target are the most popular in drug discovery as they are applicable to large protein complexes and require

low protein concentrations. A wide variety of techniques have been developed over the years.

Saturation transfer difference (STD) developed by Mayer and Meyer[80] was one of the first methods proposed for ligand observed screening of fragment libraries and remains one of the most popular. This method consists of selectively saturating a resonance of the protein. The saturation is efficiently spread over the entire protein *via* spin diffusion and if a ligand binds the saturation is propagated from the protein to the compound by cross relaxation at the ligand-protein interface. This method requires a large excess of ligand and allows measurements using a low protein concentration. STD can also be used to obtain structural information on the complex using a technique known as epitope mapping[81] (see below). Although widely used, STD does suffer from being sensitive to the size of the target (where bigger proteins are better), artifactual magnetization transfer and a limited affinity range (10 nM < K_D < 1-2mM).[82, 83] Typically, STD is used in combination with other ligand observed techniques such as waterLOGSY (see below) to eliminate false positives from screening data.[84]

One interesting modification to STD, called Group Selective-STD, directly saturates certain classes of ^{1}H's of the target (e.g. amide) in a selective manner. The method avoids the dependence on spin diffusion which is inefficient for proteins less than 20 kDa. For ^{15}N labeled targets an elegant manner of directly saturating the amide ^{1}Hs has been developed.[85] In order to insure that saturation is absolutely specific for the target, the method uses a half-filter to eliminate magnetization from ^{1}Hs not attached to ^{15}N spins. However, the required ½ J^{15N1H} delay may reduce sensitivity due to relaxation.

When the ligand resonances do not overlap with the amide region of the protein, the half-filter can be removed to improve sensitivity. Although perhaps not ideal for screening large libraries of ligands that will inevitably contain resonances near those of the amide protons of proteins, the method has been put to good use to elucidate protein-ligand structures (see below).[86]

One of the reasons for the popularity of STD is the ability to apply it to a wide variety of systems to detect ligand binding. Recent applications of STD include lipopolysaccharides,[87] glycoproteins[88] and nucleotide sugar transporters.[89] Perhaps the most interesting and challenging application has been to detect binding of ligands to membrane proteins, either purified GPCRs[90] or in live cells.[91]

Water-ligand Observed via Gradient SpectroscopY (WaterLOGSY) is a commonly used alternative to STD.[92] Based on the observation that ligand binding sites frequently also contain bound waters, the technique makes use of the abundance of water to efficiently transfer magnetization to a ligand in close proximity *via* multiple pathways. As the method is very sensitive, requires low protein and modest ligand concentrations, it has found frequent application in fragment based drug discovery. However, one has to take care that the magnetization is not transferred to the ligand *via* an artifactual mechanism such as chemical exchange and therefore, as with STD, other ligand observed techniques are frequently combined with WaterLOGSY. A nice example of this combination came from the Novartis group who used WaterLOGSY in association with T1ρ relaxation[93] to screen a small library of 500 compounds for binding to Abl kinase where the active site was blocked with imatinib.[94] Interestingly, this approach yielded biophysically validated ligands that targeted the myristate site. However, the compounds

were non-functional inhibitors. The reason for this is discussed below in the section Protein-Ligand Structures. Another recent example demonstrates the capability of the combination of virtual screening and WaterLOGSY to generate validated hits for protein-protein interaction targets.[25] The calcium binding protein S100B binds the C-terminus of p53 where it is thought to inhibit the transcription regulation function and may play a role in the progression of cancer. A collection of 123,000 commercially available compounds was screened using 3 different *in silico* approaches; importantly however, the ensemble of NMR structures of S100B was used as input for each. 280 hits from the virtual screen were purchased and binding to S100B was assayed using WaterLOGSY. The most interesting compounds from WaterLOGSY were titrated into ^{15}N-labelled protein samples to confirm binding and determine the binding site (yet another example of the use of protein observed NMR to validate screening hits). This approach successfully yielded 5 selective inhibitors that bound at the p53 binding site with reasonable ligand efficiency. The most potent inhibitor was soaked into a crystal of S100B and indeed bound to the same site as a peptide from p53.

In the original implementation of WaterLOGSY, water magnetization is destroyed prior to acquisition. Subsequently one must wait in order for equilibrium magnetization to be restored. Jahnke and coworkers recently described a polarization optimized version (PO-WaterLOGSY) in which (a portion of the) water magnetization is selectively returned to the Z axis, partially negating the need for a relaxation delay. Thus the PO-WaterLOGSY is said to speed throughput by a factor of 3-5. Aroma-WaterLOGSY, which selects that aromatic signals of ligands, effectively achieves similar results to PO-WaterLOGSY using a slightly different approach.[95]

Just as STD can provide details of the ligand-protein interaction that can help define the ligand binding mode, WaterLOGSY can provide information on the ligand-water interaction. Solvent Accessibility and protein Ligand binding studies by NMR spectroscopy (SALMON)[96] utilizes the difference in the sign of the NOE for free and protein-bound ligands to determine which portions of the ligand are solvent exposed. This information was sufficient to differentiate two possible ligand binding orientations that were compatible with the electron density derived from X-ray diffraction experiments.

Interligand NOEs for PHARmacophore MApping (INPHARMA) is a method closely related to trNOE (see below) with the twist that the NOE is relayed by the protein.[30] INPHARMA transfers magnetization between 2 small molecules that compete for binding to the same site *via* an NOE to or from the protein. INPHARMA allows indirect mapping of the binding pocket structure of a macromolecule: if the structure of one of the ligands bound to the protein is available, then an unknown ligand can be oriented on the protein. A particularly impressive example of this was the recent elucidation of the binding orientation of a number of ligands of the GPCR GPR40.[90] Although not strictly a magnetization transfer method, Interligand NOEs (ILOEs) measures direct NOEs between 2 ligands that simultaneously bind a protein and are within 5-7 Å of each other. The method can be used to screen for pairs of ligands that can be subsequently elaborated based on the relative orientation provided by the NOE information.[31,97] In the latter publication, fragment screening identified two ligands that simultaneously bound to pantothenate synthetase from *M. tuberculosi.* However, no structural information was available and the hydrophobic ligands aggregated and bound

the protein specifically and non-specifically. Chemical modification of the fragments addressed the aggregation and using the ILOE experiment the orientation of the 2 fragments was established. Subsequent linking yielded an 860 nM inhibitor from 800 μM and 1 mM fragments, suggesting that the binding of the fragments was not perturbed in the linked compound.

1.2.3. Fluorinated molecules.

^{19}F NMR has proven to be a useful tool in drug discovery. ^{19}F NMR is sensitive (the γ is similar to the ^{1}H) and the spins resonate across a broad spectrum. Although direct ^{19}F screening of fragments has been proposed,[98] the method clearly requires a fluorine atom in each member of the library and thereby limits the number and type of compounds available. However, when known ligands[99] or enzyme substrates are available,[100] then incorporation of fluorine into these molecules can lead to a high throughput, reliable assay that selects for ligands for a specific binding site (*via* competition binding) or with biochemical activity. More recently the combined use of cryoprobe technology and optimized pulse sequences has led to considerable increases in the throughput (up to 10 fold) of this experimental approach.[101] The group of Giralt has taken the approach of simultaneously screening for inhibitors of 2 different proteases by ^{19}F labeling substrates specific for each and mixing them.[41,102] In addition to throughput considerations, ^{19}F NMR can be used to characterize fragment screening hits. In one interesting case, ^{19}F labeled ATP was used to help characterize fragments binding to the kinase PDK1 and a number of allosteric activators were discovered.[43]

2. Hit prioritization

High-throughput screening (HTS) has been widely used in drug discovery to screen large libraries (up to 3 or 4 million compounds) for (typically) biological modulation of a target.[103] HTS screens may result in a few hundred to many thousands of positives. Among these "hits" many are not the result of specific, reversible interaction with the target but rather non-specific effects such as interference with the assay itself or commonly aggregation.[19] Thus, there is an emerging need to triage the list of HTS hits. At Pfizer, Miller and colleagues have conducted hit prioritization campaigns using biochemical and biophysical techniques.[104] The first validation step consists of an orthogonal enzymatic assay to validate the results of the HTS screen. Then, to ensure integrity, a resynthesis step of the compound is carried out followed by SAR studies. Finally, STD and ITC are used in parallel as a last step of validation. This extensive triage of compounds allows classification in distinct series and the majority of unwanted compounds are eliminated in the early stages. One caveat is that in our experience, many HTS hits have very limited aqueous solubility rendering ligand observed NMR challenging.

As hits from a fragment or HTS screen are progressed, the affinity typically drops into the nM range. In this range k_{off} becomes slow on the NMR timescale and the K_D is less than the protein concentration typically required for either ligand or protein observed NMR. In such cases determining the affinity using NMR is problematic. However, in a series of compounds often the relative affinities are more important than the absolute

affinities. Jahnke and colleagues have developed a method to acquire the relative affinity of 2 compounds whose absolute affinities may be close or widely separated.[47] The method can be implemented using either protein or ligand observed spectroscopy and has proven useful in the elaboration of fragments. Although the protein observation method is the most robust, it is more restrictive than the ligand observed.

3. Protein-Ligand structures

Structure based drug design has become the *de facto* standard method for targets in which high throughput crystallography is enabled. An analysis of FBDD projects suggested that the success rate (as measured by achieving inhibition better than 100 nM) was 3 fold higher when 3D structures of protein-ligand complexes was available.[105] NMR and X-ray crystallography remain the only methods for protein structure determination, while, for commercial drug discovery, it is imperative that this information is rapidly available. There have been a number of exciting developments that are beginning to enable NMR to provide structural information in the time frame demanded. Below, we point out a few that are, in our opinion, the most exciting.

While ultimately it is ideal to have high resolution 3D structures of the protein-ligand complex, information about the conformation of the ligand itself can be valuable. The transferred NOE (trNOE) is a simple technique that can be applied to a variety of targets. A NOESY spectrum of a ligand in rapid exchange with the target is acquired under conditions of ligand excess. NOEs between ^{1}Hs of the ligand only build up when bound to the larger protein. The intra-ligand NOE information can then be used to

determine the conformation of the ligand in the bound state. This method is applicable to large proteins of which only small amounts are available, but requires relatively high compound concentrations and therefore the compounds must be quite soluble. TrNOE has been used to determine the conformation of the bound form of weak inhibitors of MurD ligase, an enzyme that contributes to the formation of peptidoglycan.[21] Epitope mapping was carried out using STD and the NMR information was used to restrain molecular dynamics calculations generating models of the protein-ligand complex. Although crystal structures of the complex were available, they did not explain the weak protein-ligand interaction. The NMR experiments suggested that the ligands are highly dynamic in the bound state, potentially explaining the unexpectedly low potency.

If the protein can be isotopically labeled, group selective STD can provide information on the types of ^{1}Hs on the protein that are close to ^{1}Hs of the ligand. This approach has been elegantly used to map the carbohydrate binding site of galectin 1.[86] Protein amide resonances can be directly and efficiently saturated, while ligand resonance can be eliminated, if necessary, through a half filter. However, due to the relative size of the STD signal and the natural abundance of ^{13}C, carbon attached ^{1}Hs must be selectively saturated to avoid artifacts. In principle, if the structure of the target is known, restraints derived from such a study could be used to guide molecular docking efforts even in the absence of the sequential assignment.

The chemical shift is a readily available parameter that contains information on not only the chemical environment, but also the conformation about a nucleus. As noted, perturbations to the NMR spectrum of a protein have long been used for detecting binding of e.g. small molecules. Returning to the work on Abl kinase, differences in the

chemical shift of Val 525 were correlated with the presence of myristate in its binding pocket and the attendant activation of the enzyme.[94] By selectively ^{15}N labeling Val residues of Abl, the authors developed a conformation specific assay that could be used to detect allosteric ligands and differentiate agonists from antagonists. The assay was used to screen hits for their ability to induce the same active conformation as myristate. This assay confirmed the activity of the known allosteric modulator, GNF-2. However, GNF-2 had liabilities related to non-specificity. The assay was then used to discover novel allosteric agonists of Abl.

Unfortunately, despite numerous efforts, it is not yet possible to directly convert chemical shift information into structure constraints. However, there have been significant achievements in using protein CSPs to guide docking. In one example, amide CSPs from an [^{15}N,^{1}H] HSQC were used to guide the docking of compounds to the β subunit of the transcription factor CFB.[106] Thirty five compounds had been selected from *in silico* screening amongst which, four were validated in the HSQC experiment. After directed library synthesis, two, more potent compounds were selected for detailed experimentally guided docking. The docking suggested two possible orientations for the cores of the two compounds, one of which gave a better correlation to the experimental data. The binding site was distal to the binding site of the 2nd subunit of CFB, Runx1, and the compounds acted as allosteric inhibitors. In an attempt to step beyond the state of the art, González-Ruiz and Gohlke have developed a method that quantitatively exploits amide proton CSPs for protein-ligand docking.[107] This method combines standard scoring by the DrugScore function[108] (which describes the protein –ligand interactions) with scoring ligand poses with respect to their agreement with experimental CSP data. The

comparison is achieved by back calculating the CSP based exclusively on ring currents. After testing the approach on a set of crystal structures, the authors applied it to 3 real world cases for which CSP data was available. In two of the 3 cases the docking pose was quite good, however the 3rd differed substantially. The authors' analysis of the reasons for the differences highlight the complexities of the approach, which include the effects of hydrogen bonds between the ligand and protein and CSPs resulting from conformational and/or dynamic changes that occur upon ligand binding. In yet another example, both CSPs and intermolecular NOEs were used to guide the docking of the prodrug Losartan to glycoprotein VI, a target for anti-thrombotics.[109]

Instead of trying to derive direct structural constraints from CSPs, one can analyze the data in terms of binding models. Auto-FACE (Auto-Fast chemical Exchange analyzer) fits titration data from an e.g. [^{15}N,^{1}H] HSQC experiment to multiple possible models of protein-ligand interaction.[110] The models include simple two site binding, multiple site with or without sequential binding and allosteric contributions. The authors claim that residues directly involved in ligand binding can be differentiated from those affected by e.g. conformational changes, by analyzing the "initial rate of perturbation". Isothermal titration calorimetry suggested a 4 state binding model for the system being investigated where the last state was non-specific. The NMR titration data was analyzed in terms of the binding constant and the initial rate and magnitude of the perturbation. This data suggested two distinct binding sites with different affinities for the compound, which agreed well with the ITC data (free, low affinity and high affinity complex). Interestingly, the two sites correspond to two sites predicted by JSURF[111] using CSPs from low and high concentrations of the ligand. For a more extensive review of the use of

chemical shift information for protein structure elucidation see the recent comprehensive review by Mulder and Filatov.[112]

Solving the NMR structure of a large protein remains a challenge. Recently, Schwalbe and colleagues achieved specific resonance assignment of roughly half of the primary structure of DXR (DOXP reductoisomerase), an 87 kDa homodimer that is a potential anti-infective target,[113] by using 3D heteronuclear experiments with uniform ^{15}N,^{13}C and ^{2}H labeling. The authors measured intermolecular NOEs between the protein and the cofactor NADPH thus potentially enabling the structure of the complex to be determined using the known crystal structure. However, no intermolecular NOEs to a bound inhibitor could be detected. Selective labeling of amino acid residues can be a powerful alternative to uniform labeling that can enable sequential assignment of even very large proteins. In one particular example, the 723 residue Malate Synthase was labeled uniformly along the backbone while the sidechains of I, L and V residues were selectively labeled with a linear pattern of ^{13}C,^{2}H where the terminal methyl groups were additionally protonated.[114] Such a scheme can enable through bond assignment of the backbone and nearly complete assignment of the methyl resonances. Since methyl containing residues are typically well distributed at ligand binding sites,[65] the method of Tugarinov and coworkers could be used to determine structures of even very large ligand-protein complexes in an efficient manner. Nonetheless, through bond assignment techniques are insensitive and time demanding for large proteins. Direct assignment of selectively labeled methyl resonances could make the approach highly efficient. Very recently, the group of Clore[115] has used paramagnetic relaxation enhancement (PRE) to achieve direct, stereo-specific assignment of methyl resonances of the 27 kDa N-terminal

domain of the *E. coli* protein Enzyme 1. Five different site specific Cys mutants of the protein were used to attach nitroxide radical containing tags. The distant dependent PRE rate was measured for each methyl using a simple HMQC experiment and the information was used in a Metropolis MonteCarlo calculation to determine the assignments. The PRE assignments were then used to compare experimental *vs* predicted methyl-methyl NOEs for validation purposes. While the technique clearly requires the generation of multiple cysteine mutants (the authors suggest 1 mutant per 6 kDa), when feasible it should lead to rapid resonance assignment.

In many cases the structure of a protein may be known but obtaining crystals with a ligand bound, particularly the weak binding ligands typical of fragment based drug discovery, may not be possible. Paramagnetic tags provide one attractive approach to solving this issue. The group of Otting[116] first showed that it is possible to calculate the structure of ligand in rapid exchange with a protein to which a lanthanide ion is bound at an intrinsic metal binding site. If the orientation of the paramagnetic tensor is known with respect to the protein coordinates, then both distance and angular information can be derived from the magnitude of the pseudocontact shift. A more recent example of the approach applied to a protein in which a lanthanide binding tag has been fused to the protein was provided in the section on paramagnetism for ligand discovery.[78] The additional example of a non-covalently bound peptide tag was also provided.[77] These methods are particularly exciting as once the tagged protein is available and the tensor orientation determined, structure constraints are rapidly determined by titrating the ligand or tag and acquiring simple 1D ^{1}H spectra of the ligand. In principle then, the methods should be quite high throughput.

While X-ray crystallography remains the structural method of choice when available, it may have shortcomings. A recent example is provided by the structure of the growth factor receptor-bound protein 2 (Grb2) SH2 domain.[117] The crystal structure consisted of a domain swapped dimer to which inhibitors bound to a hinge region between the dimers. The use of perdeuterated protein and assignment of the intermolecular NOEs gave a well defined solution structure which consisted of a 1:1 complex of protein and inhibitor. The correct structure suggested vectors for elaboration/improvement of this non-phosphate containing ligand. The protein Mesencephalic astrocyte-derived neurotrophic factor (MANF) provides a second interesting example of a significant difference between the crystal and solution structure.[58] Where the crystal structure indicated a disordered C-terminal domain, the NMR structure clearly demonstrated the existence of a SAP domain. Interestingly, the SAP domain of the protein Ku70 inhibits the pro-apoptotic activity of Bax. A similar role for the SAP domain of MANF would be consistent with the demonstrated ability of MANF to protect neurons of rats. When a crystal structure is not sufficient to observe loops due to high disorder, NMR can become an asset, especially when important residues for activity are located in these loops. In one example, two inhibitors of the proline rich kinase (PYK2) were thought to stabilize the relatively rare DFG-out, inactive form of the kinase.[118] As with most kinases, the DFG loop was not visible in the crystal structure of the protein-ligand complex. ^{15}N labeling of Phe residues was used in conjunction with TROSY-HSQC to demonstrate changes in the spectrum consistent with the expected rigidification of the DFG loop upon stabilization in the out conformation.

4. In-cell NMR spectroscopy

In-cell NMR spectroscopy holds great potential as a tool to discover and validate the interaction of small molecules with pharmaceutical targets in a natural milieu. We make no attempt to comprehensively review this field here (we refer the reader to a recent review on the subject),[119] but rather focus on two recent developments that we feel are important for drug discovery.

STINT-NMR was developed as a method to study protein-protein interactions inside living bacterial cells using NMR spectroscopy.[120] This method uses sequential expression of the two proteins whose interaction is to be studied. The target protein is first expressed using ^{15}N labeled medium, in order to obtain a [^{15}N,^{1}H]-HSQC spectrum inside the living cell. The growth medium is subsequently changed and the interacting protein is overexpressed without labeling. The HSQC spectrum of the target changes upon increasing concentration of the interacting protein. Screening small molecule interactor libraries (SMILI-NMR) uses STINT-NMR to screen libraries of small molecules for protein-protein interaction inhibitors with activity in the cell.[121] SMILI-NMR represents the first effort to use in-cell methods to screen compound libraries. Combined with the methods described below this could prove to be a powerful method to combine biophysical sensitivity with biological activity.

While the initial in-cell NMR experiments used isotope labeled proteins in *E. coli*,[122] for drug discovery it is critical to assess small molecule functions in eukaryotic and preferably mammalian cells. While the seminal work of Selenko[123] demonstrated that in-cell NMR in eukaryotic cells was feasible, this work relied on injection of purified,

isotope labeled protein into Xenopus oocytes. More recently the elegant work of Inomata and colleagues has demonstrated high resolution HSQC spectra of proteins inside living human cells. This method uses cell penetrating peptides[124] to deliver protein into the cytosol. The peptides are covalently linked to the protein and the release is made upon enzymatic activity or reductive cleavage.[125] Importantly, the method was used to show both FK506 and rapamycin binding to ^{15}N labeled FKBP in HeLa cells.

5. Perspectives

In 2008 a group of experienced NMR spectroscopists in industry and academia wrote on the perspective for the use of NMR for drug discovery.[126] It is greatly exciting to note that many of the developments that have been highlighted in this review are in areas that the authors of the perspective defined as being important for the future expansion of NMR within industry. Although NMR now faces increased competition from other methods in the area of ligand discovery, the value of its contributions throughout the drug discovery pipeline make the investment in instrumentation and expertise well worthwhile.

Chapter 3

Functional immobilization of EphA4 and application of fragment based drug discovery to identify inhibitors

Protein kinases are considered to be one the most important group of drug target and aberrant kinase activity is implicated in a variety of human diseases. Recent studies have demonstrated that a particular receptor tyrosine kinase, EphA4, is overexpressed in several cancers. The ability to modulate the activity of this enzyme is therefore an attractive therapeutic strategy and we aimed to find small molecules inhibitors of EphA4 activity. We sought to apply a ligand screening technology, Target immobilized NMR Screening (TINS), to this particular target. The kinase domain of EphA4 was functionally immobilized on resin and a library of drug fragments was subsequently screened using the TINS technology. Hits obtained with TINS were biochemically characterized using a commercially available enzyme inhibition assay. In addition, selected compounds were submitted to crystallographic structure determination.

Introduction:

Eph receptors regulate numerous developmental and cellular processes such as cell attraction/repulsion, adhesion/detachment and migration, thereby influencing morphogenesis and organogenesis.[127] Recently, EphA4 has been identified as a validated target for the treatment of certain skin lymphomas.[58] Subsequently, others have demonstrated the importance of EphA4 in prostrate cancer and pancreatic cancer.[128, 129] Therefore we aimed to develop small molecules inhibitors targeting the kinase domain of EphA4. The kinase domain of EphA4 is centrally located in the cytoplasmic portion of the protein, adjacent to a sterile α domain (SAM) and a C terminal PDZ domain binding motif[130] (Fig. 3.1). N-terminal to the kinase domain is a short amino acid sequence referred to as the juxtamembrane segment (JMS). The JMS domain contains tyrosine residues that must be phosphorylated to allow full activation of the kinase activity,[131] a mechanism of regulation that is unique to the Eph receptors. Extensive studies of this domain have shown that the JMS acts as an additional level of regulation and directly influences catalytic activity.[132, 133]

Fragment Based drug discovery (FBDD) has become a powerful approach for generating novel chemical entities targeting pharmaceutical targets. FBDD screening libraries consist of a small set (1,000-20,000 compounds) of molecules which are called "fragments" i.e. their physicochemical properties comply with the "Rule of Three".[29] (Ro3, Mr < 300 Da, cLogP <3, H-bond donors < 3, H-bond acceptors < 3, number of rotatable bonds < 3 and TPSA (total polar surface area) < 60 $Å^2$). These fragments typically bind to the target with K_D greater than 10 μM. To detect these weak binding

affinities, sensitive biophysical techniques are required, such as Target Immobilized NMR Screening (TINS) that has been developed in our laboratory.[39, 40] In this technique, the target protein to be screened is immobilized on a solid support (a chromatography resin). The immobilized target, along with a sample of a reference protein, is placed inside a flow injection, dual-cell sample holder[40] in the NMR magnet. The compounds to be screened are injected in presolubilized mixes containing 3-8 compounds at 500 μM each. For each mix, an independent 1D ^{1}H spectrum of the compounds in the presence of the target or a reference protein is acquired using spatially selective spectroscopy.[134] Compounds binding to the target are detected by the decrease in peak amplitude which results from the greatly enhanced transverse relaxation experienced in the bound state. Weak, non-specific interactions typically observed between compounds and proteins are cancelled by the presence of a reference sample. This reference sample consists routinely of the PH domain of the cellular kinase AKT as this protein was shown to be essentially refractory to small-molecule binding.[135] The aim of this study was to functionally immobilize the tyrosine kinase domain of EphA4, in order to submit this protein by Target Immobilized NMR Screening (TINS). We investigated several different immobilization chemistries resulting in covalent or non-covalent attachment leading to either oriented or non-oriented samples. Immobilization on resin was quantified and the amount of active protein on resin was determined. As functional immobilization of this protein was not as straightforward as expected, several TINS screens were performed. Using a commercially available enzyme inhibition assay, the hits obtained with TINS were biochemically characterized and interesting compounds were submitted to crystallographic structure determination.

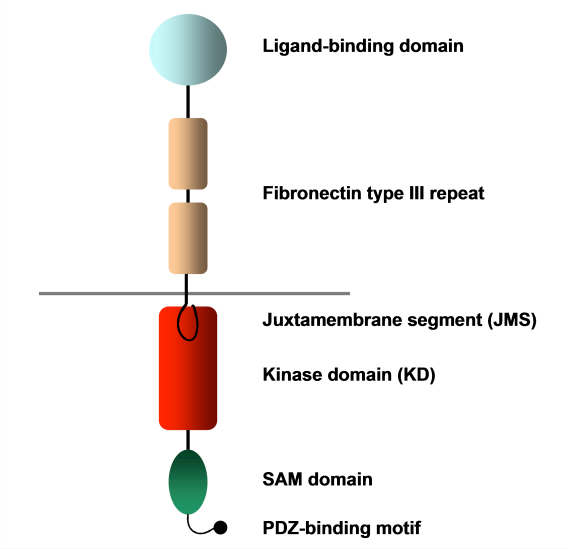


Figure 3.1: Domain structure of Eph receptors. The extracellular domain consists of an N-terminal ligand binding domain, followed by a cysteine-rich domain and two fibronectin type III repeats[136, 137] a single membrane-spanning segment, a juxtamembrane segment (JMS), a tyrosine kinase domain (KD), a sterile α domain (SAM) domain and a C-terminal PDZ domain binding motif.

Material and methods

Material:

The Actigel ALD resin (Sterogene, Carlsbad, CA, USA) was used in a 50% slurry in all experiments. The resin was washed three times with filtered water then twice with PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4, 137 mM NaCl, 2.7 mM KCl) by centrifugation at 500 rpm for 5 min and resuspended in an equal volume of the same buffer.

SNAP-tag Kinase Domain:

Expression and purification of the SNAP-tagged Kinase Domain:

The murine EphA4 kinase domain (amino acids 606-846) was cloned into a pSNAP-tag®(T7) Vector from New England Biolabs (MA, USA) generating an N-terminal SNAP-tag fused to the kinase domain: SNAP-KD with an intervening region containing an amino acid sequence efficiently hydrolyzed by the tobacco etch virus (TEV) protease. The cloning procedure was performed by colleagues at the LUMC. The expression plasmid was transformed into *E. coli* BL21-AI competent cells. Protein expression was induced overnight at 18°C with 0.2% arabinose. A cell pellet corresponding to 1 liter of culture was resuspended in 30 ml of lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM DTT and 1 mM EDTA) supplemented with a protease inhibitor cocktail (complete

EDTA-free) from Roche, lysed using a French press and centrifuged at 35,000 x g for 45 min at 4°C.

SNAP-KD Immobilization:

The benzylguanine (BG) resin was prepared by coupling a benzylguanine derivative BG-polyethyleneglycol (PEG)-NH_2 (New England Biolabs (MA, USA)) to Actigel ALD resin. BG-PEG-NH_2 was first solubilized in N,N-dimethyl formamide (DMF) to a final concentration of 100 mM, then diluted in water to a final concentration of 5 mM. One ml of Actigel ALD resin, the solution of BG-PEG-NH_2 and coupling reagent (1 M $NaCNBH_3$) at a final concentration of 0.1 M were mixed together and placed at 18°C and gently rotated for 18h. After immobilization, the supernatant was removed after a 5 min centrifugation at 500 rpm. To block the remainder of the free aldehyde sites on the resin, the resin was resuspended with 1 ml Tris buffer (100 mM Tris, pH 7.5) and 0.1 M of coupling reagent and incubated for 2 h at room temperature. The reducing agent was then removed by washing the resin four times with Tris buffer (100 mM Tris, pH 7.5).

Crude lysate of SNAP-KD was incubated with 1 ml of benzyl-guanidine resin and placed at 4°C with gentle rotation for 18h. Then the resin was washed with phosphate buffer. The SNAP-KD resin was then incubated with 20 mM ATP and 20 mM $MgCl_2$ in order to fully activate it.

To obtain the isolated KD in solution, the immobilized SNAP-KD was treated with TEV protease (1:1 molar ratio) fused to a 6His tag expressed and purified according to published protocols[138] for 24 hr at 20°C in TEV activity buffer (25 mM NaPi, 125 mM NaCl, pH 7.4, 5 mM DTT). The resin was centrifuged for 5 min at 500 rpm, and the

supernatant containing the KD cleaved and TEV protease was collected. To remove TEV protease, the sample buffer was exchanged for binding buffer and then applied to a 5 ml HisTrap column.

Quantification of SNAP-KD immobilization efficiency:

To determine the approximate amount of SNAP-KD immobilized, a method of protein quantification using Coomassie Blue R was employed[139]. Protein standards were made with a known concentration of an immobilized reference protein (AKT).

A portion of the resin suspension to be assayed was placed in a microcentrifuge tube (40 µl of 50% slurry). The resin was incubated with 100 µl of the Staining Solution (0.25% Coomassie Blue R, 30% methanol, and 10% acetic acid) for 10 min at room temperature. The sample was then centrifuged at 13000 rpm for 1 min and the supernatant was removed. The pelleted resin was washed twice with 1 ml of the Destaining Solution (30% methanol and 7.5% acetic acid). After a few seconds of mixing to insure that the pellet was well suspended in the solution, the sample was centrifuged for 1 min at 13000 rpm. The bound dye was eluted by suspending the pellet in 0.4 ml of a solution containing 1% SDS and 1% sodium bicarbonate. To this suspension was added 0.6 ml of methanol and the solution was mixed well to ensure thorough suspension of the resin. The resin was pelleted by centrifugation at 13000 rpm for 3 min and the supernatant fluid was decanted and saved for determination of the bound protein content of the resin sample. The amount of dye in the supernatant fluids from the elution step was determined by measurement of the absorbance at 595 nm.

Quantification of SNAP-KD activity:

To determine whether the activity of the unbound protein was comparable to the one immobilized on resin, we used a filter binding assay.[140] Each reaction consisted of 3 μM of EphA4 in solution or immobilized in PBS buffer. Aliquots of 40 μl were placed in microcentrifuge tube and reaction solution consisting of 20 mM $MgCl_2$, 10 μM poly E4:Y1 (Sigma, St. Louis, MO), 50 μM ATP + 0.1 μCi [γ-32P]-ATP per reaction (Perkin Elmer, 6000 Ci/mmol 10 mCi/ml) in assay buffer (60 mM HEPES, pH 7.5, 0.2 mM DTT, 10 mM KCl) was added to the samples. After one hour incubation at room temperature with gentle shaking, 10 μl aliquots were spotted on P81 phosphocellulose chromatography paper (Whatman, cation exchanger) and allowed to air-dry. The P81 filter arrays were washed three times with 300 ml of 75 mM orthophosphoric acid, for 10 min each time with gentle shaking. Washing progress was monitored with a Geiger radioactivity counter. The P81 filter arrays were finally washed briefly with acetone and left to dry under a chemical hood. In order to quantify the quantity of radioactivity incorporated, different dilutions of [γ-^{32}P]-ATP were spotted on the P81 phosphocellulose chromatography paper and left to dry under a chemical hood. The P81 paper arrays were exposed to the storage phosphor screens for an hour then analyzed with a BioRad phosphoimager. Signals were quantified and the background was subtracted.

Avi fusion Kinase Domain

Construction of the Avi fusion Kinase Domain plasmid:

All restriction endonucleases, T4 Polynucleotide Kinase and T4 Polynucleotide Kinase Reaction Buffer were purchased from New England Biolabs (MA, USA).

The pET28a vector plasmid (Invitrogen) containing the murine EphA4 kinase domain described in Chapter 4 was used as a template for this cloning procedure. The vector was digested with NheI-HF and BamHI in appropriate buffer for one hour at 37°C. The reaction mixture was run on a 1% agarose gel and then gel purified with GFX gel purification kit (Qiagen).

The single stranded oligonucleotides (oligonucleotide 1: 5'-CTA-GCG-GCC-TGA-ACG-ATA-TTT-TTG-AAG-CGC-AGA-AAA-TTG-AAT-GGC-ATG-AAG-3' and oligonucleotide 2: 5'-GAT-CCT-TCA-TGC-CAT-TCA-ATT-TTC-TGC-GCT-TCA-AAA-ATA-TCG-TTC-AGG-CCG-3') were purchased from Eurogentec and reconstituted in DNase/RNase free water at a concentration of 200 µM. To increase the ligation efficiency, 500 pmol of oligonucleotides were 5' phosphorylated (5' PO_4) with 10 units of T4 Polynucleotide Kinase in presence of 400 µM ATP and 5 mM DTT in T4 Polynucleotide Kinase buffer for 1 hour at 37°C.

An equal amount of phosphorylated oligonucleotides were mixed together in a PCR tube and heated at 95°C for 2 min, and then re-annealing was done at RT for 5 min.

Ligation reactions were performed with a molar ratio of linearized vector to re-annealed oligonucleotides of approximately 1:6 for 18 h at 16°C in the presence of T4 DNA ligase in a final volume of 10 µL. T4 DNA ligase was heat-inactivated (65°C for 20

min) before transformation into *E. coli* DH5α cells. Transformed cells were propagated for 1 h before spreading onto Luria-Bertani (LB) agar. Plasmid DNA was purified from cultivated positive colonies, and analyzed by PCR reaction with oligonucleotide 2 and the T7 Forward primer (Eurogentec). The PCR reaction consisted of one cycle of 1 minute of denaturation at 95°C, 30 cycles of 1 minute of denaturation at 95°C, 1 minutes of hybridization at 55°C and 2 minutes of elongation at 72°C and a last cycle of 1 minute of denaturation at 95°C, 1 minutes of hybridization at 55°C and 5 minutes of elongation at 72°C. Plasmids giving the expected PCR product were submitted to confirmation via sequencing.

Expression and purification of the Avi-KD:

The Avi fusion EphA4 kinase domain expression plasmid was transformed into *E. coli* BL21-CodonPlus (DE3)- RP competent cells. Protein expression was induced overnight at 18 °C. A cell pellet corresponding to 1 liter of culture was resuspended in 30 ml of lysis buffer (50 mM Sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% Tween 20, 10% glycerol and 0.01% beta-mercapto-ethanol) supplemented with a protease inhibitor cocktail (complete EDTA-free) from Roche, lysed using a French press and centrifuged at 37,000 x g for 45 min at 4°C. The supernatant was applied to a 5 ml HisTrap HP column (GE Healthcare) with a peristaltic pump at 4°C for 1 hour (Binding buffer: 20 mM Tris-HCl pH 7.6, 500 mM NaCl and 5 mM imidazole). After extensive washing, the recombinant protein was eluted with binding buffer plus 500 mM imidazole. The sample buffer was then exchanged for 10 mM Tris pH 8.0, 500 mM Potassium glutamate. The biotinylation reaction was performed according to manufacturer

recommendations (Avidity, LLC). The protein sample at a concentration of 40 μM was incubated with 2.5 μg birA enzyme in 50 mM bicine buffer, pH 8.3 supplemented with 10 mM ATP, 10 mM MgOAc, 50 μM d-biotin. After 30 min incubation at 30°C, the reaction mixture was applied to a 1 ml HisTrap HP column (GE Healthcare), and washed with binding buffer. The biotinylated protein was eluted with binding buffer plus 500 mM imidazole.

Immobilization of biotinylated KD

The streptavidin resin was prepared by coupling streptavidin (Invitrogen) to Actigel ALD resin. Streptavidin was dissolved in 1.2 ml PBS at a concentration of 150 µM. The solution of Streptavidin, 1.2 ml of Actigel ALD resin and coupling reagent (1 M $NaCNBH_3$) at a final concentration of 0.1 M were mixed together and placed at 4°C and gently rotated for 18h. After immobilization, the supernatant was removed after a 5 min centrifugation at 500 rpm. To block the remainder of the free aldehyde sites on the resin, the resin was resuspended with 1 ml Tris buffer (100 mM Tris, pH 7.5) and 0.1 M of coupling reagent and incubated for 1h at room temperature. The reducing agent was then removed by washing the resin four times with Tris buffer (100 mM Tris, pH 7.5).

The biotinylated kinase domain of EphA4 was obtained as described above and was incubated with 1 ml of streptavidin resin and placed at 4°C with gentle rotation for 18h. The resin was then washed with phosphate buffer.

Target Immobilized NMR Screening

Immobilized EphA4 and the reference protein were each packed into a separate cell of a dual-cell sample holder[40] in deuterated PBS buffer. A homemade packing reservoir has been built to fit on top of the dual-cell sample holder and double the volume of each cell. The resin (as a 50% slurry) is pipetted in to each cell one at the time, allowed to settle by gravity and packed at a pressure of 0.5 bar. The cell was attached to a Spark autosampler via capillary tubing and inserted into an 8 mm, ^{1}H selective, flow-injection probe in a 500 MHz magnet. Mixes of typically 3-4 fragments were made by dilution of a 100 mM stock of each compound in d_6-DMSO such that the final compound concentration was 500 μM. Upon injection of each mix into the dual-cell sample holder, the flow was stopped and spatially selective Hadamard spectroscopy[26] was used to acquire a 1D ^{1}H spectrum of each sample separately in the EphA4 cell and in the reference cell. A CPMG T_2 filter of 2 ms or 80 ms depending on the screen was added prior to signal acquisition to suppress resonances arising from the resin. Periodic injection of a mix containing reference molecules (500 μM Imidazole, 500 μM Na Acetate, 250 μM TSP, and 500 μM TMA) during the entire screen was used as a measure for reliability of sample injection.

Kinase Inhibition Assay

Kinase activity was measured using the ADP Hunter Plus assay kit according to the manufacturers recommendations (DiscoveRx, Birmingham, UK). In this assay ATP

turnover, and the subsequent generation of ADP, is measured in a two-step process. Reaction mixtures (20 μl) were prepared in black polypropylene 384 well flat bottom plates (Greiner, Stonehouse, UK) and contained 100 nM His-tagged EphA4-KD, 33 μM ATP (determined as Km for this kinase preparation), 200 μg/mL E4:Y1 (Sigma, St. Louis, MO) as substrate and increasing concentrations of inhibitors. Assay buffer (pH 7.4) contains 20 mM NaCl, 15 mM HEPES, 20 mM $MgCl_2$, 1 mM EGTA and 0.1% Tween20. After 75 min at 30ºC, reagents A and B from the kit were added and the fluorescence intensity was measured with a Tecan infinite F200 Reader (Tecan, Grödig, Austria) after 60 min incubation at 30ºC, with an Ex of 535 nm and Em of 590 nm and 25 reads per well. Standard curves, ranging from 30 nM to 75 μM ADP (dynamic range of the kit), were used as an internal assay control. Background was measured in reactions containing no substrate and subtracted from those containing E4:Y1. For confirmed hits, the effects of the inhibitors on the enzymes in the ADP detection step were determined in assay buffer and reagents A and B only.

The percentage of inhibition (PIN) per reaction was calculated as follows:

$$PIN\ (\%) = [1- (A_{fragment}-A_{positive\ control})/(A_{negative\ control}-A_{positive\ control})]\ x100\%$$

where $A_{fragment}$ and $A_{negative\ control}$ are the signal in Response Units (RU) measured in the presence of the fragment and without fragment, respectively, whereas $A_{positive\ control}$ is the signal in RU in the presence of dasatinib, a known inhibitor that abrogates EphA4 enzymatic activity completely[47, 141]. The assay was performed in triplicate.

Fragment soaking

Crystals of the kinase domain of EphA4 were grown as described in Chapter 4. Stocks solutions of compounds in DMSO were diluted with mother liquor and added to EphA4 crystal (final concentration from 1mM to 30 mM) for an exposure time ranging from 5 min to 36 hrs at 20°C. Then crystals were flash-frozen using 20% glycerol plus the reservoir solution. Diffraction data was collected at 100K at the ESRF (Grenoble, France). Datasets were merged and processed using XDS[142] and Pointless[143] from the CCP4[144] suite and the structure was solved by molecular replacement using AMoRe[145]. The apo EphA4 kinase domain (PDB ID: 2Y6M) was used as a search model (*c.f.* Chapter 4). The structure was subsequently refined using PHENIX[146] and COOT[147] was used for model building.

Results

Functional immobilization of EphA4

In order to insure the physiological relevance of ligands discovered using TINS, it was important to retain full enzymatic activity of EphA4 upon immobilization. Therefore a number of immobilizations chemistries were investigated including: covalent, non-oriented immobilization via Schiff's base chemistry, oriented, covalent immobilization via glutathione S-transferase (GST), oriented covalent immobilization via the reaction of the SNAP tag with benzyl guanine and oriented non-covalent immobilization via biotin-streptavidin binding.

The first trials aimed to immobilize the kinase domain via the Juxtamembrane segment JMS-KD (amino acids 591-846). The JMS-KD protein was expressed and purified according to previous publications.[132] The purified protein was immobilized on sepharose resin using non-oriented immobilization via Schiff's base chemistry between the primary amines of the protein and the aldehyde groups present on the commercially available resin. The immobilization was carried out in the presence of AMP-PNP, an analog of ATP, to protect labile amines in the active site. The protein was immobilized at a concentration of 100 μM on resin and the activity of the protein was monitored by incorporation of radiolabeled phosphate from [γ-^{32}P]-ATP into a peptide substrate (Fig. 3.2A). We observed that the amount of ^{32}P incorporated into the peptide substrate, and hence the activity of the protein, is influenced by the immobilization. Immobilized protein retained only 20% of the activity of an equivalent amount of protein in solution.

Moreover, the presence of AMP-PNP did not appear to influence the enzymatic activity of immobilized protein. Therefore, we investigated other immobilization chemistries.

As the protein as purified is a fusion with GST, it was logical to assay immobilization via covalent interaction with glutathione. The GST fusion JMS-KD protein was immobilized on glutathione beads from GE Healthcare at a concentration of 80 μM and the enzymatic activity was assessed. As shown in Figure 3.2B, the protein immobilized on resin presented activity comparable to the protein in solution. Since immobilization via the GST tag appeared to yield fully functional protein, it was decided to carry out a TINS ligand screening experiment (see below). However, the results gave an unexpected profile: most of the compounds showed preferential binding to the reference. For that reason we decided to investigate other approaches to functionally immobilize EphA4.

As immobilization of the wild type JMS-KD was unsuccessful, other protein constructs from the literature were investigated. Previous mutational studies of the tyrosine kinase of EphA4 indicated that mutation of tyrosine 742 to alanine in the kinase domain yielded a constitutively, highly active protein.[133] This mutant protein (JMS-KD Y742A) was expressed and purified. Immobilization of JMS-KD Y742A on resin using non-oriented immobilization via Schiff's base chemistry in the presence of AMP PNP was assayed as described above. As shown in Figure 3.2C, the mutant JMS-KD Y742A has more activity than the wild type protein in solution. However, when immobilized on resin the mutant protein retained only 5% of the activity of an equivalent amount of protein in solution.

Since various attempts to immobilize the JMS-KD protein failed to yield sufficiently active enzyme, it was decided to focus on the kinase domain only (amino acids 606-846). For this purpose a second oriented immobilization based on a fusion to the SNAP-tag protein was employed.[148] The SNAP-tag is a protein based on human O6-alkylguanine-DNA-alkyltransferase (hAGT) and can be covalently attached to beads conjugated to benzyl guanine. The SNAP tag protein was fused to the Kinase Domain (SNAP-KD) with a TEV protease recognition site in between. The amount of immobilized protein was quantified on resin and the activity was assessed and compared to the KD in solution. Since purification of the SNAP-KD was based on the covalent bonding between SNAP and benzyl guanine, we could not obtain pure SNAP-KD in solution. Therefore the soluble KD, obtained by proteolytic digestion of the immobilized, purified SNAP-KD, was used for enzymatic activity comparison (Fig. 3.2D). The soluble KD cleaved from SNAP exhibited greater enzymatic activity than the JMS-KD, as previously reported[132]. The immobilized SNAP-KD has similar enzymatic activity to the KD in solution (Fig 3.2D). Consequently, the SNAP-KD protein was screened using the TINS technology.

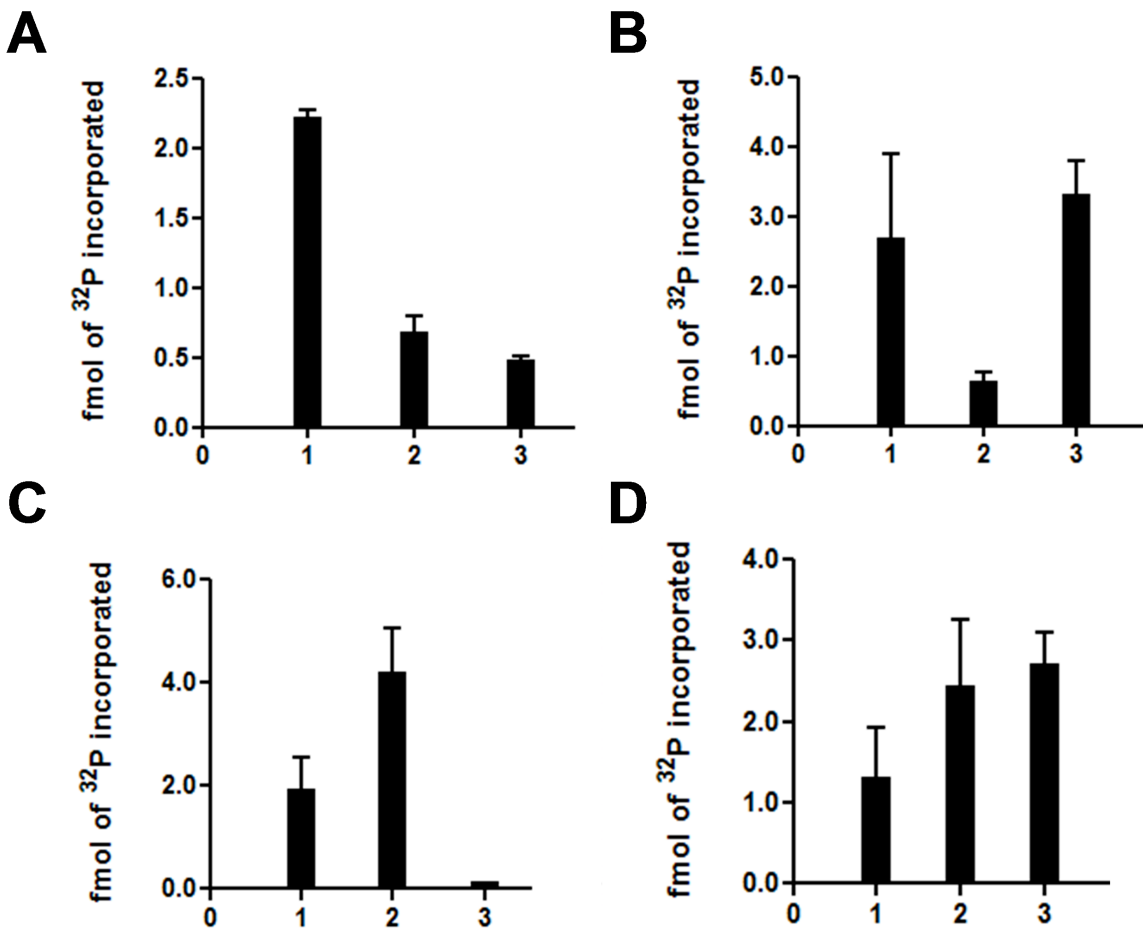


Figure 3.2: Enzymatic activity of EphA4 immobilized using a variety of approaches. The amount of ^{32}P incorporated into the peptide substrate is presented. **A)** Non-oriented immobilization via Schiff's base chemistry. (Column 1) JMS- KD in solution. (Column 2) JMS- KD immobilized without AMP-PNP. (Column 3) JMS- KD immobilized in presence of 5 fold molar excess of AMP-PNP. **B)** Oriented, covalent immobilization via GST fusion. (Column 1) GST tagged JMS-KD in solution (Column 2) JMS-KD immobilized on Actigel via Schiff based chemistry (Column 3) GST tagged JMS-KD immobilized on glutathione beads. **C)** Non-oriented immobilization via Schiff's base chemistry. (1) JMS-KD in solution (2) JMS-KD Y742A in solution (3) JMS-KD Y742A immobilized on Actigel ALD resin in presence of 5 molar excess of AMP PNP. **D)** Oriented covalent immobilization via SNAP fusion. (1) JMS-KD in solution (2) KD in solution (3) SNAP-KD immobilized on benzyl guanidine resin. The activity in each bar is normalized for the amount of protein.

Target Immobilized NMR Screen (TINS)

Ligand screening using TINS requires immobilized protein. During the search for functional immobilization, the GST fusion JMS-KD proved to be functional when immobilized on glutathione beads. We decided to pursue ligand screening using this protein and as a reference we selected the GST protein immobilized on glutathione beads. A complete screen of about 1200 compounds was performed. The data resulting from the screen could be analyzed directly by comparing peaks from TINS spectra with the individual reference spectra of each fragment (Fig. 3.3). The ready identification of binders enabled a totally automated analysis of the complete screening data from Fourier Transform and phasing to comparison of peak amplitudes and ligand identification. The analysis was performed using in-house written routines implemented in TopSpin, the spectrometer control and data analysis software.

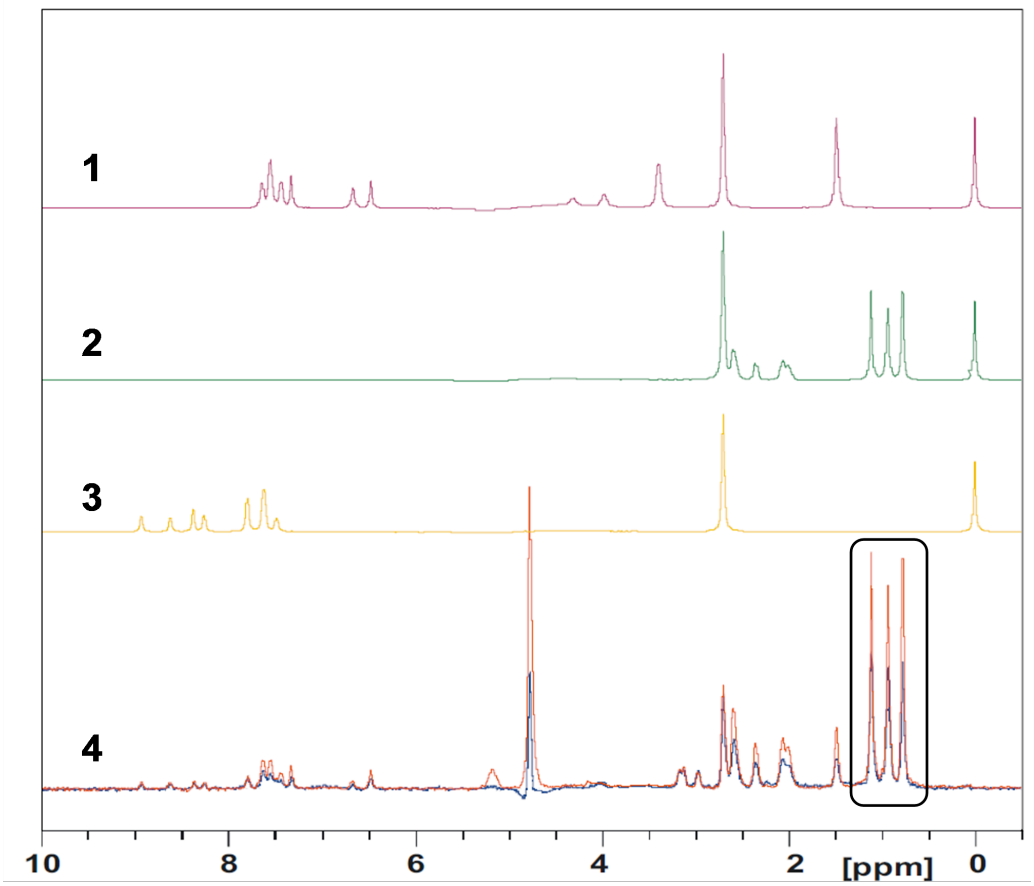


Figure 3.3: Detection of ligand binding to immobilized EphA4 using TINS. The 1D ^{1}H NMR spectrum of 3 different fragments in solution (1-3) is shown. The ^{1}H NMR spectra of a mix of 3 fragments in presence of EphA4 (4- blue spectrum) or in presence of reference (4- red spectrum) are presented. The encircled region shows reduction in peak amplitudes that are expected upon specific binding of a fragment to the immobilized target.

The results of the screen are presented in Figure 3.4A. Binding is represented by the peak heights in the presence of the target (T) divided by peak heights in the presence of the reference (R). A T/R ratio higher than 1 indicates preferential binding of a compound to the reference and conversely, a ratio lower than 1 shows preferential binding to the target protein. The frequency of occurrence of the T/R ratio of each compound gives a profile of the screen that is unique for each target. Surprisingly, in the present screen most of the compounds showed preferential binding to the reference: only 6 compounds exhibited binding to the target. Apparently the isolated GST protein has high levels of non-specific small molecule binding.

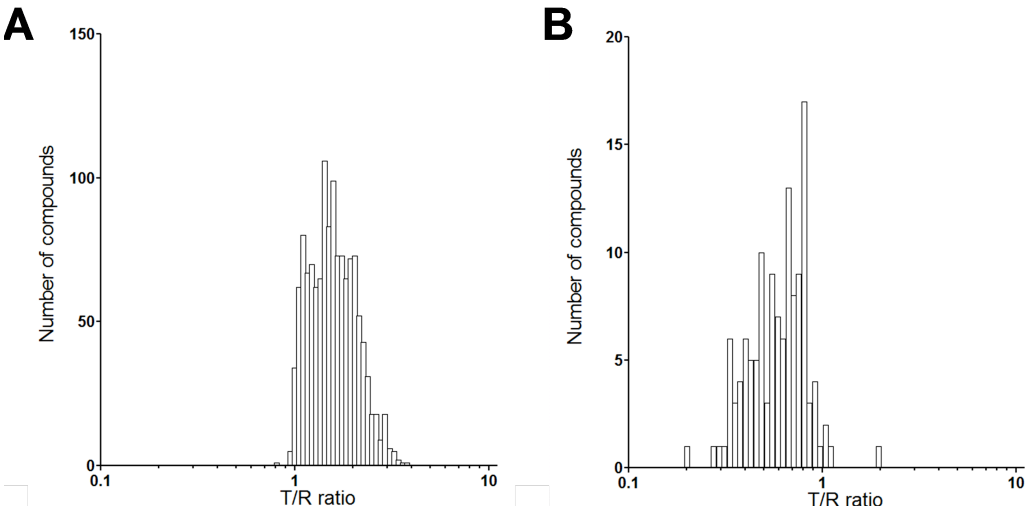


Figure 3.4: T/R distribution for TINS ligand screening by compound. The graph presents the number of compounds with a T/R in each of the bins presented on the x-axis. T/R ratios are binned into equal groupings from 0.1 to 1 and 1 to 10, the number of compounds in each bin is plotted. A) Screen of GST fusion JMS-KD, B) Screen of SNAP-KD using SNAP as a reference protein.

Since the screen of the GST immobilized KD did not yield the desired fragment ligands, we investigated an alternative approach. As shown in Figure 3.2 immobilization of the SNAP-KD fusion via SNAP resulted in an apparently fully functional protein. Therefore, we used this immobilization procedure to screen the fragment library a second time. As seen in the screen of GST-KD, the choice of the reference is crucial. In the screen of the SNAP-KD, the SNAP protein alone was initially immobilized as a reference. As shown in Figure 3.4B, out of the 650 compounds screened most exhibited preferential binding towards the target. Despite the fact that SNAP had proven to have very low intrinsic small molecule binding capacity[149], it is not an ideal reference. We then decided to use the PH domain of human AKT as a reference protein as it had been used in a large number of successful TINS screens. The TINS ligand screen of SNAP-KD vs. AKT PH was performed using two different T_2 relaxation periods of 80 ms and 2 ms. The relaxation period is used to suppress the broad resin signal but it also allows more time for the fragments to diffuse and encounter an immobilized binding site. A short T_2 time will result in detection of compounds biased toward high affinity while a long T_2 time will detect compounds biased towards specificity for the target protein (*cf.* supplementary Figure S1 A and B)

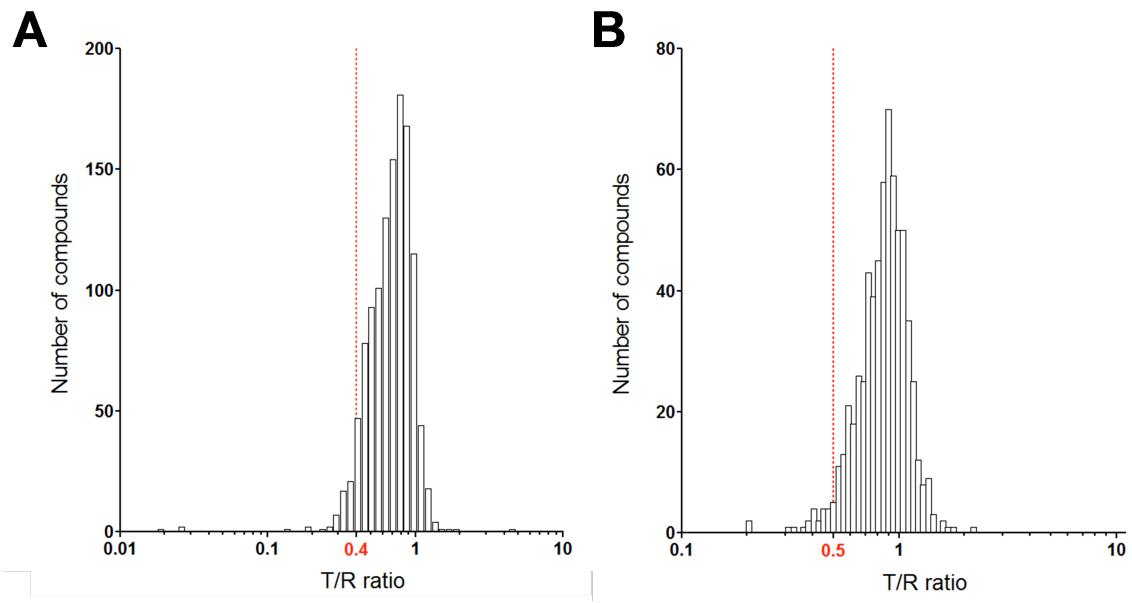


Figure 3.5: Ligand screen of the SNAP fusion KD vs. AKT as a reference: T/R distribution by compound. The graph presents the number of compounds with a T/R in each of the bins presented on the x-axis. T/R ratios have been binned into equal grouping from 0.1 to 1 and 1 to 10, the number of compounds in each bin is plotted. A) Screen with a T_2 (relaxation period) of 80 ms, B) Screen with a T_2 (relaxation period) of 2 ms.

Once the screens were completed, in-house developed routines were used to transform and analyze the data. The T/R ratios are binned into equal groupings from 0.1 to 1 and 1 to 10 and the number of compound in each bin is plotted (Fig. 3.5). The cut-off, although arbitrary, was selected in order to differentiate compounds with better target binding properties from the rest of the fragments. With a T_2 delay of 80 ms, 1191 compounds were screened resulting in 82 "hits" (positives) for SNAP-KD, defined as fragments which had a T/R ratio lower than 0.4. The screen using a T_2 delay of 2 ms was run on 651 compounds and resulted in 26 hits with a T/R ratio lower than 0.5. The resulting hit rate for SNAP-KD with T_2 of 80 ms and 2 ms was respectively 6.8% and

3.9% which is well within the range observed with TINS (3-9.5%). Hits from both screens were selected and then further characterized using an enzyme inhibition assay.

Hit validation using enzymatic assay

The TINS assay identifies compounds that simply bind to EphA4. A logical next step is to characterize the biological activity of the TINS hits using an enzyme inhibition assay. The enzyme inhibition study was performed at a single concentration of each fragment to generate an initial ranking of the biological activity. Hits from the TINS screens (at 2 ms and 80 ms), as well as several non-binders, were selected for further characterization (Table 3.1).

Table 3.1: Characterization of TINS fragments screening hits via enzyme inhibition.

	TINS screen at 80 ms		TINS screen at 2 ms	Non-binders	Total
	T/R <0.4	0.4<T/R<0.5			
Compounds assayed	81	46	35	19	182
Interfering compounds	4	0	5	3	12
Non active compounds	60	43	27	14	144
Inhibitors	15	3	3	2	23
Activators	2	0	0	0	2

A total of 182 compounds were individually assayed at 500 μM in the kinase inhibition assay. Fragments were considered "inhibitory" when they exhibited a percentage of inhibition or PIN value above 10% which was 3 times the average standard deviation of the entire enzymatic assay. Compounds which interfered with the assay were excluded from the analysis. Out of the remaining 169 compounds, 25 exhibited biological activity: 23 inhibitors and 2 activators. The 23 fragments inhibitors were further analyzed for potency (IC_{50}) by dose response experiments (Fig. 3.6, Table 3.2 and Appendix B). A complete description of the fragments studied in this thesis is provided in Appendix A.

Dose-response experiments were carried out with increasing fragment concentration, from 8 nM to 1.5 mM, while ATP and peptide substrate were kept in excess. Dasatinib, a known inhibitor of Eph family members including EphA4,[47, 141] was used as a positive control. Dasatinib titrated over 2 log orders and exhibited a Hill coefficient of -2.9. The Hill coefficient is a measure of cooperativity in a binding process and for enzyme targets with simple kinetic mechanisms (no complex allosteric regulation, for example), a Hill slope of one is expected. However, in the present situation, as the enzyme concentration significantly exceeds the K_D value of Dasatinib, a steep dose-response (i.e. a Hill slope > 1) is expected.[150]

Intriguingly, most of the dose-response curves for the fragments tested appeared to saturate at 50% of full inhibition at high compound concentration (~1.5 mM). Four of the 23 fragments showed less than 20% of full inhibition and/or steeper than expected Hill coefficients (Appendix A). The most potent compounds were **1** and **2**, which showed saturation at 50% and 75% respectively of full inhibition. Compounds which had an IC_{50} < 750 μM were prioritized for hit optimization (see below).

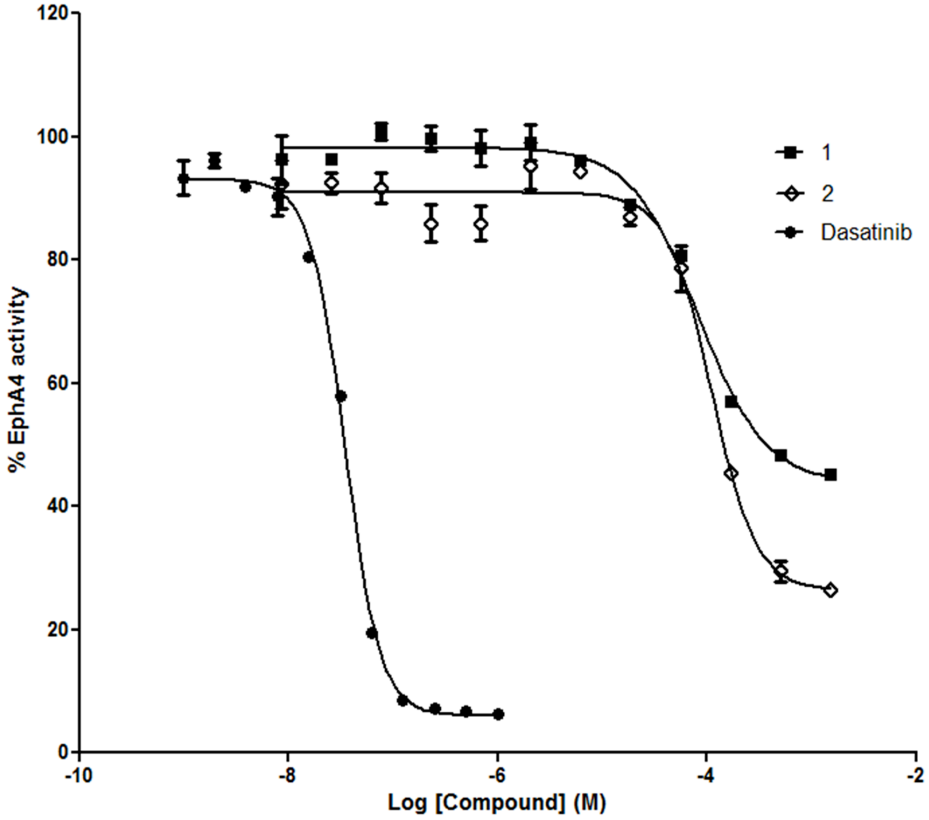


Figure 3.6: Inhibition curves of selected hits from the TINS screen. The curve represents the mean ± Standard Error of Mean (S.E.M) of experiments performed in triplicate. All curves were from GraphPad Prism using a sigmoidal dose response curve model with variable slope.

Table 3.2: Inhibition values of selected hits from the TINS screen. Inhibitory constants, top and bottom plateau values and Hill coefficient are provided.

Compound	IC_{50} (μM)	Y value at the top plateau the curve	Y value at the bottom plateau of the curve	Hill coefficient
1	82 ±20	105.3±0.9	46.9±2.3	-1.40±0.18
2	110±26	107.0±1.2	30.9±2.9	-2.04±0.33
Dasatinib	0.024±0.001	98.8±0.9	8.1±0.8	-2.91±0.17

Crystallography

To evolve the low affinity TINS fragments into high affinity inhibitors, compounds **1**, **2**, **5**, **6** and **16** to **20** were submitted to hit optimization via *in silico* methods. In this approach fragments are progressively grown by means of modeling and docking to make further interactions with the protein. Optimized compounds were then resubmitted to validation via the enzymatic assay. Then, to determine the binding mode, we attempted to determine the structure of the most potent compounds bound to EphA4 KD by crystallography.

Many strategies are available for obtaining crystals of protein-ligand complexes.[151] The most resource-effective method of obtaining the structure of a protein–ligand complex is by soaking the ligand of interest into ligand-free protein crystals (*i.e.* incubation of the crystal with the compound). As the crystal structure of the kinase domain of EphA4 was solved (*c.f.* Chapter 4), soaking experiments were set up in order to obtain crystals of protein-ligand complexes. The structure of a compound from the *in silico* screening effort bound to EphA4 KD was successfully solved using this soaking procedure (*c.f.* Chapter 5). We have therefore initially set up soaking experiments for the fragment screening hits using similar conditions i.e. 30 mM fragment concentration with incubation times of 6 hrs. Various fragment concentrations (from 1 to 30 mM) and exposure time to the crystal (from few minutes to 36 hours) were assayed without any observable degradation (monitored using a microscope) of the crystal. Most of the compounds did not display any or only minor solubility problems (as judged by a lack of visible precipitate) under these conditions. In total 35 crystals containing 13 different

compounds either optimized *in silico* or TINS fragments were sent to a synchrotron in order to obtain high resolution diffraction data. Most of the crystals exhibited weak diffraction i.e. less than 7 Å, however the datasets for compounds **1**, **2**, **6**, **17**, as well as a dataset for an optimized fragment originated from **2** were of sufficient quality to solve using molecular replacement techniques. Unfortunately, no electron density was observed for any of these compounds which could have various explanations. First, that under these conditions the ligands did not bind to the KD or there was insufficient occupancy within the crystal roster. Second, it is possible that the fragments were not sufficiently ordered to yield observable electron density or that the fragments bind in multiple ways to the same site.

As the pH of the crystallographic conditions (pH 5.5) was quite different from the pH of the kinase assay (pH 7.5), there was a possibility that this pH difference might influence fragment binding. Thus we decided to use biophysical techniques as an additional filter for fragment binding before submitting them to crystallographic analysis. Multiple biophysical techniques, such as 2D NMR and Surface Plasmon Resonance (SPR) were employed at a pH similar to the crystallographic conditions to further characterize fragment binding to the EphA4 KD. This biophysical analysis of the TINS fragments is presented in Chapter 6.

Functionality of Biotinylated EphA4

During the SPR analysis of TINS fragment screening hits, the SNAP tagged KD protein yielded poor results (*c.f.* Chapter 6 for details). These SPR results combined with

the surprisingly low confirmation rate of TINS screening hits in the enzyme inhibition assay led us to doubt the validity of the initial screen on the SNAP tag fusion protein. Therefore we decided to pursue a third immobilization procedure in order to determine whether the SNAP fusion had unknowingly influenced the ligand screening procedure. In this immobilization approach, streptavidin is first immobilized on beads using Schiff's base chemistry and subsequently used to bind the biotinylated KD. In order to biotinylate the EphA4 outside of the KD (and minimize the possibility of interference with enzymatic activity), a 15 amino acid peptide (GLNDIFEAQKIEWHE), called an Avi Tag, was inserted at the N-terminus. This sequence is efficiently biotinylated by the biotin ligase (BirA) from *E. coli*.[152]

The biotinylated KD was immobilized on the streptavidin resin and the functionality of the protein was compared to protein in solution *via* a spectrophotometric coupled assay.[131] We observed a steady rate of NADH oxidation of 2.66 (ΔA_{340}/s)/μmol of kinase and 1.77 (ΔA_{340}/s)/μmol of kinase for soluble and immobilized protein respectively (Fig. 3.8). The rate of the reaction was close to the value previously reported of 2.12 (ΔA_{340}/s)/μmol of kinase.[131] These results suggest the biotinylated protein remained active when immobilized.

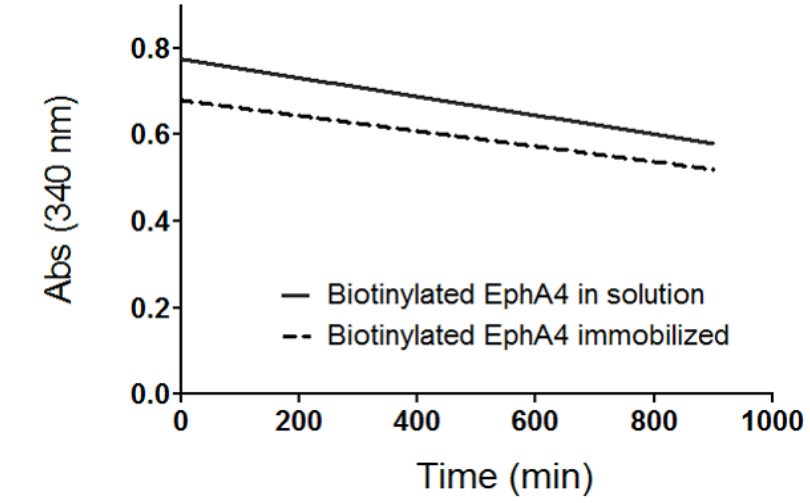


Figure 3.7: Enzymatic activity of biotinylated EphA4 immobilized and in solution. Analysis was performed using a coupled in vitro kinase assay where production of ADP is coupled to the oxidation of NADH through pyruvate kinase and lactate dehydrogenase. Both proteins were assayed at 1 μM. The graph shows the reduction of NADH observed by a decrease in absorption at 340 nm.

Target Immobilized NMR Screen (TINS) of biotinylated EphA4

In order to screen the ZoBio fragment collection, biotinylated KD and biotin were immobilized on the streptavidin resin at a solution equivalent of 100 μM. A total of 1306 fragments were screened as described above with a T_2 delay of 2 ms. The screen resulted in 62 hits for biotinylated EphA4, defined as fragments with a T/R ratio < than 0.8. The profile of this screen is summarized in Figure 3.8. For more clarity we will refer to biotinylated KD screen as Biot-KD screen. Orthogonal validation of the screening hits was generated using SPR (*c.f.* Chapter 6).

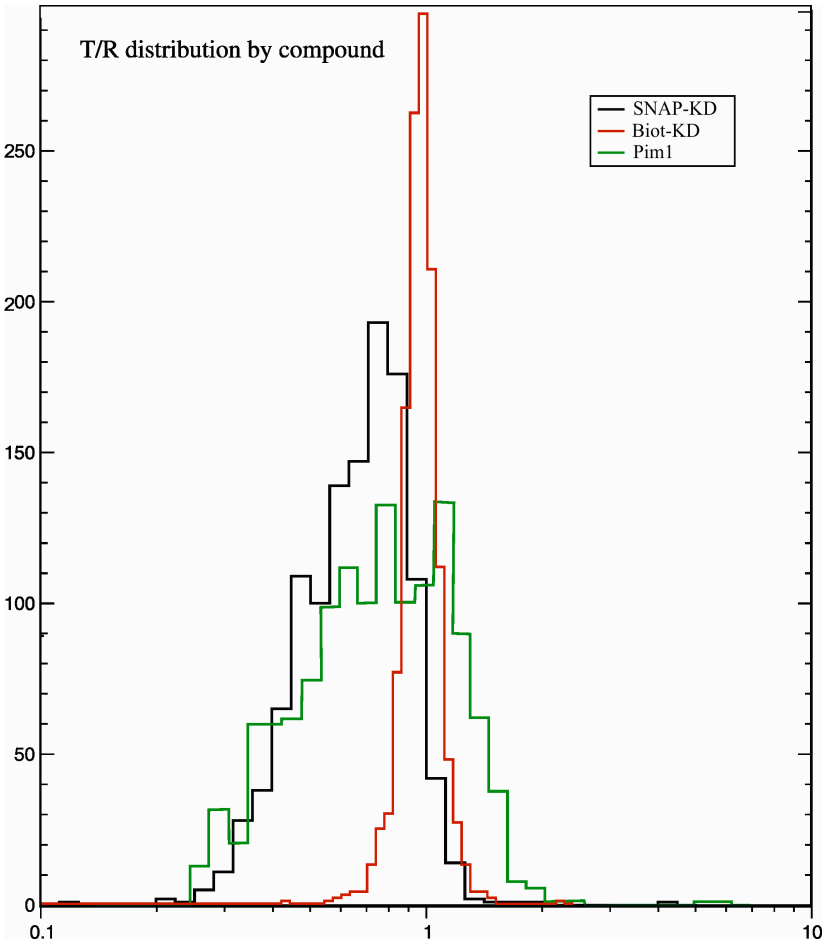


Figure 3.8: Comparison of TINS ligand screens: T/R distribution by compound. The graphs present the number of compounds with a T/R in each of the bins presented on the x-axis. The T/R ratios were binned into equal grouping from 0.1 to 1 and 1 to 10. The black histogram is from the TINS screen of SNAP EphA4 KD while the screen of biotinylated EphA4 KD is represented in red. For comparison, the TINS profile of an unrelated kinase, Pim1, is shown in green.

In order to judge the influence of the immobilization method we compared the results of the screens of the SNAP and biotinylated EphA4 KDs (Fig 3.8). The profile of the Biot-KD screen is different from the profile for the 2 ms screen of SNAP-KD. The peak of the Bio-KD profile is centered on 1 where in the SNAP-KD screen it is shifted

toward target binding. A priori, one would expect that a large collection of randomly chosen small molecules would not show bulk preferentially binding to a protein if non-specific interactions are properly accounted for. The data suggests that non-specific binding is better cancelled out in the Biot-KD screen and therefore we hoped that hits from this screen would fare better in subsequent analysis.

In addition to comparing the profile of the Biot and SNAP-KD screens, it can also be informative to compare the T/R value for individual compounds. As shown (Fig. 3.9A, comparison of T/R values acquired using the same NMR parameters yields no correlation). In Figure 3.9B, shows the T/R values for the subset of compounds from both screens for which enzyme inhibition data are available. Of the compounds selected as "hits" in the Biot-KD screen, 5 were characterized in the enzyme inhibition assay. Two out of 5 inhibited EphA4 activity, meanwhile none of the 4 "hits" from the SNAP-KD exhibited inhibition. Although the numbers are small, the limited data does suggest that the Biot-KD screen yielded slightly more biologically active compounds. Out of the 430 compounds tested in both, only one is considered a "hit" in both screens (T/R ratio lower than 0.5 in the SNAP-KD 2 ms screen and T/R ratio lower than 0.8 in the Biot-KD screen) , yet this compound did not inhibit enzymatic activity.

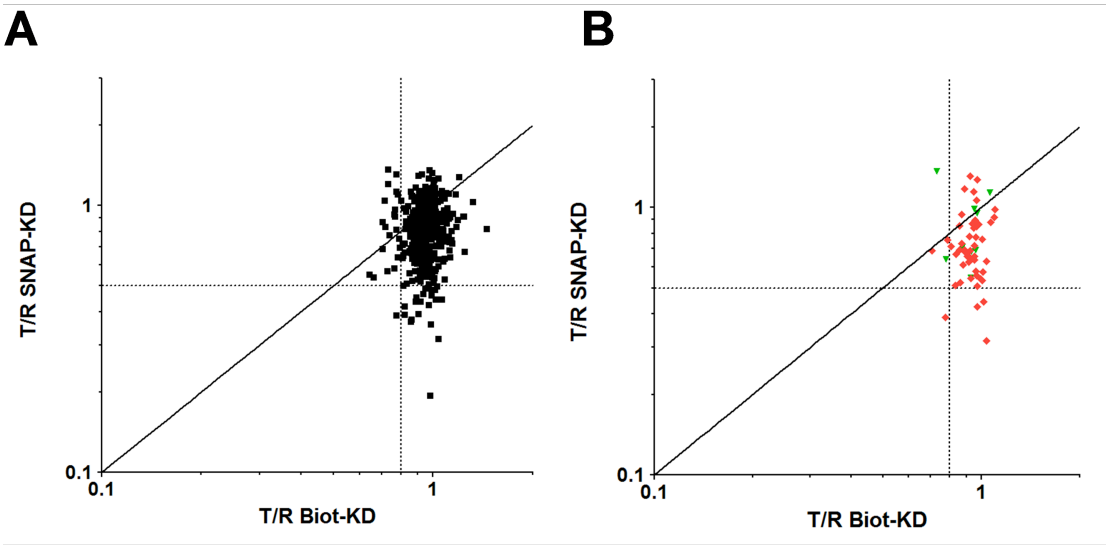


Figure 3.9: Comparison of the KD TINS screens at a T_2 relaxation time of 2 ms. In x axis is presented the T/R ratio of the biotinylated KD screen (Biot-KD) is plotted and in y axis the T/R ratio of the SNAP-KD screen is shown**.** A) The 430 compounds in common to both screens are plotted. B) Only the compounds for which enzyme inhibition data was available are shown. Compounds represented by green dots were inhibitory, while those in red dots were inactive.

Discussion

The goal of this study was to find hit matter that could potentially be developed into high affinity inhibitors of the kinase activity of EphA4. In order to accomplish this goal we screened a fragment library for binding to EphA4 using TINS. This method requires immobilization of the protein with retention of function. Achieving a high level of functionality of immobilized EphA4 required investigation of a number of immobilization chemistries. We first attempted to immobilize the protein via Schiff's base chemistry. This approach, based on the direct, covalent attachment of protein via primary amines, has proven useful in numerous other TINS screens including kinases.[39] A previous direct immobilization of the protein kinase Pim1[149] was performed in the presence of AMP-PNP and yielded functional protein. However, in our case this direct method did not yield functional immobilized EphA4. Even in the presence of AMP-PNP to protect the binding site, the protein remains mostly inactive on resin. At pH 7.4, the Schiff's base reaction is rather specific for the amino terminus. However, lysines with perturbed pKa's may also react at this pH. One such residue could be Lys^{656} whose importance for enzyme activity has been established in mutagenesis studies.[131] Lys^{656} is located outside the binding site, thus presence of AMP-PNP would not prevent coupling to this amino acid residue.

As the direct immobilization of EphA4 was unsuccessful, a number of other immobilization techniques were tested. First, the GST fusion JMS-KD was covalently immobilized using glutathione beads. This immobilization appeared to yield fully functional protein and subsequently the protein was screened using Target Immobilized

NMR Screening. The results of the screen showed an unexpected profile: fragments were preferentially binding to the reference protein (GST) suggesting that GST has high intrinsic (possibly non-specific) small molecule binding and therefore interferes with ligand screening. In order to overcome the difficulties encountered with the GST immobilization, the immobilization of EphA4 mutants was investigated. Previous mutagenesis studies have identified constitutively active mutants of EphA4 kinase activity that could prove useful for ligand screening.[131] Immobilization of several of these mutants was investigated but did not yield a functional immobilization of the protein.

As several other protein kinases were successfully screened using the TINS technology, we tried to understand the reason for the failure with EphA4. The major difference between these protein kinases and EphA4 resides in the presence of the JMS. As it was possible that this segment was interfering with the functionality of the immobilized protein, we decided to focus on the kinase domain of EphA4. Oriented functional immobilization of EphA4 kinase domain was achieved via a SNAP tag and the protein was screened using the TINS technology. Initially, the protein reference used was the SNAP protein itself but all the compounds screened appeared to preferentially bind to the target protein. Although SNAP has been successfully used as a reference in TINS in order to cancel weak, non-specific interactions that are observed between compound and protein,[149] in this case the reference did not seem to function appropriately. Therefore, we used the PH domain of AKT as a reference and screened nearly 1200 compounds using a T_2 delay (relaxation time) of 80 ms resulting in 82 hits. A second screen of 651 compounds using a T_2 of 2 ms yielded 26 hits. To validate these results, the TINS profile of the screen with EphA4 was compared to the one obtained with another kinase, Pim1.

As seen in Figure 3.8, the two TINS profiles were very similar, suggesting that SNAP-KD protein was behaving as expected.

Since TINS simply identifies compounds that bind to a protein, it is important to validate the hits in terms of biological activity. The hits from the TINS screen were biochemically characterized using a kinase assay which yielded a surprisingly low validation rate (15% of the fragments exhibited biological activity). For comparison, nearly 75% of hits from Pim1 showed enzyme inhibition (Masakazu Kobayashi, personal communication). Among the fragments that were characterized as inhibitory, most presented peculiar dose response curves with a steep Hill slope. One possible explanation for this observation is that the compounds precipitated at the higher concentrations in the assay conditions.[153] Moreover, the dose response-curves appeared to level off below 100% inhibition which can be characteristic of compounds that non-specifically aggregate.[150] Still, it is unlikely that all these compounds precipitated or aggregated as the library was carefully designed and the compounds assessed for solubility at high concentration (~ 3 mM).[62] Compounds that exhibited some inhibition and presented an interesting chemical scaffold were selected for crystallographic structure determination. Soaking experiments were undertaken and the complex of a compound that originated from the *in silico* screen was successfully solved (Chapter 5). However, none of the fragments that originated from the TINS screen successfully yielded a structure. The appearance of crystals was monitored with a microscope and despite the lack of visually detectable changes, most of these crystals did not diffract to high resolution. However, the experimental electron density map could be obtained for several crystals, but no density was observed for the ligand. It is possible that the fragments were not sufficiently

ordered to yield observable electron density or that the fragment did not bind the protein under the crystal formation conditions. As the pH of the crystallographic conditions is far from the pH of the biochemical assay, there was a possibility that the pH difference could negatively influence binding of the fragments. Thus we decided to employ extra biophysical filters such as SPR and 2D NMR prior to submitting compounds to soaking trials (Chapter 6).

The lack of biochemical validation of the hits from TINS and the low yield of crystal structures led us to consider another TINS screen. Moreover, SPR studies presented in Chapter 6 yielded an alternative means to achieve functional immobilization of EphA4, i.e. via the streptavidin/biotin interaction. Biotinylated EphA4 was subsequently immobilized for TINS studies. Nearly 1300 compounds were screened and the results were compared to those from the SNAP-KD screen. Surprisingly, little correlation was observed between the two screens, even if both proteins remained active when immobilized. Moreover, the TINS profiles are different for the two screens (Fig. 3.8). At this point it is not clear where the variability comes from. However, the experience within the laboratory and ZoBio is that TINS screens typically exhibit good reproducibility (Gregg Siegal, personal communication). In addition, the ability to consistently crystallize different batches of apo KD suggests that variability in the protein preparations was minimal.

Previously, Pim1 was successfully screened using TINS. Approximately 75% of the hits were validated using a biochemical assay. In addition, the X-ray crystallographic structure was determined for 37 out of 50 compounds. In our case, only 15% of the hits selected by TINS exhibited activity and we could not obtain structural information. To

gain further insight into the possible source of these problems, the compounds were analyzed with orthogonal biophysical methods including SPR and protein observed NMR as described in Chapter 6.

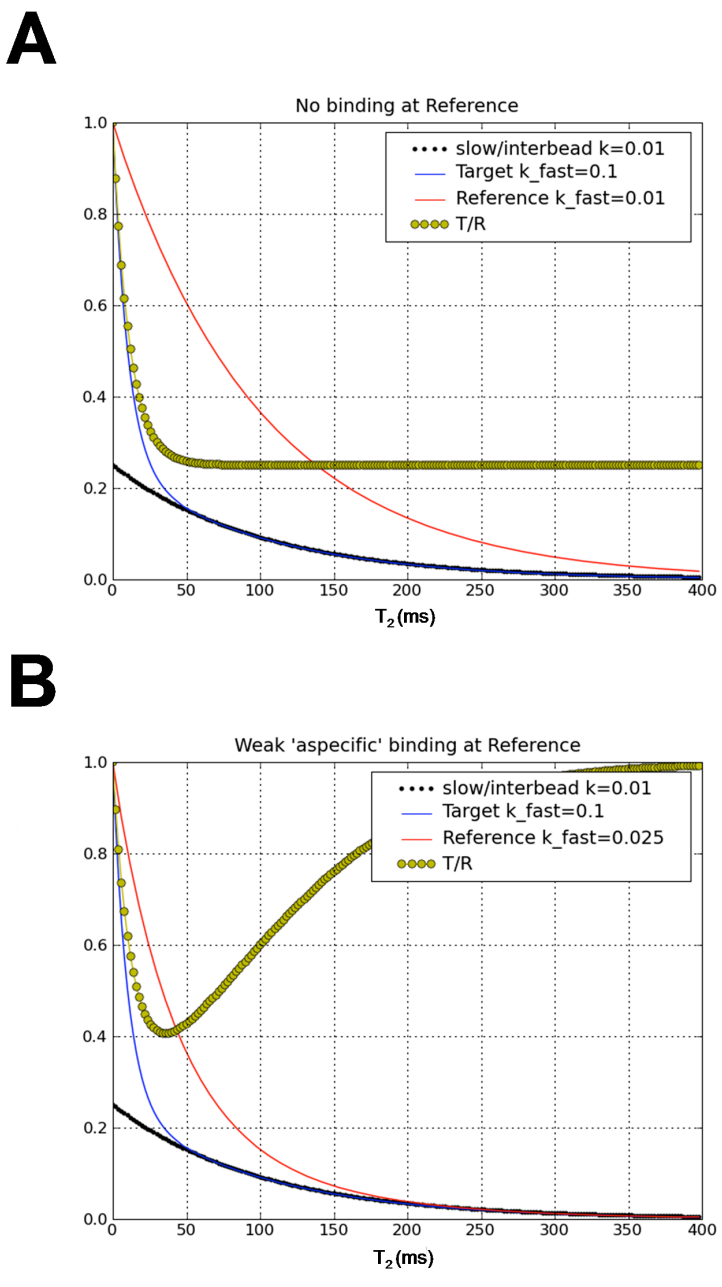


Figure S1: Simulation of the peak amplitude in the TINS experiment. The relative peak amplitude is plotted against the T_2 time in ms. The relative peak amplitude for the target and the reference is shown in blue and red respectively. The resulting T/R ratio is represented with yellow dots and the signal from the interbead space *i.e.* in solution is shown in black dots. In the simulation in A, the compound exhibits tight binding to the target and weak binding to the reference. In B, the compound exhibits the same binding to the target while binding to the reference has been simulated with only 10 fold weaker affinity. In A, the T/R ratio is indicative of binding at all T_2 times where in B binding is only detected for shorter T_2 times.

Chapter 4

Crystal structure of EphA4 kinase domain in the apo and Dasatinib bound form

The Eph family of receptor tyrosine kinases regulates diverse cellular processes while the over-expression of a member of this family, EphA4, has been reported in a variety of malignant carcinomas. To gain insight into molecular mechanisms and to facilitate structure-based inhibitor design, we solved the crystal structure of the native EphA4 kinase domain in both the apo and dasatinib bound forms. Analysis of the two structures provides insight into structural features of inhibitor binding and revealed a hydrophobic back-pocket in the ATP- binding site of EphA4 which was previously unidentified. The structures suggest a route towards development of novel and specific inhibitors.

C. Farenc, P. H. Celie, C. P. Tensen, I. J. de Esch and G. Siegal. Crystal structure of the EphA4 protein tyrosine kinase domain in the apo- and dasatinib-bound state. *FEBS letters*, 2011, **585,** 3593-3599.

Introduction

Receptor Tyrosine Kinase signaling generally contributes to a malignant phenotype by aberrantly transducing growth-promoting stimuli.[44] Both overexpression and constitutive activity of these receptors due to mutation have been demonstrated to occur frequently in various malignancies.[43] Recently, EphA4 was identified as a validated target for the treatment of certain skin lymphomas.[58] Subsequently, others have demonstrated the importance of EphA4 in prostate cancer.[59]

Eph receptors share a common topology where the extracellular region consists of: an N-terminal ephrin binding domain, a cysteine-rich region and two fibronectin type III repeats near the single membrane-spanning segment. The cytoplasmic region contains a short amino acid sequence referred to as the juxtamembrane segment (JMS), the highly conserved tyrosine kinase domain (KD), a C terminal sterile α domain (SAM) and a PDZ domain binding motif.[130] The JMS domain contains tyrosine residues that must be phosphorylated to allow full activation of the kinase activity[131] which is unique to the Eph receptors. In the crystal structure of EphB2, the JMS forms an α-helix that interacts with the kinase domain and prevents it from adopting an ordered, active structure, thus providing another level of regulation.[132] NMR studies of EphB2 have shown that upon activating phosphorylation events, the JMS and KD undergo increased conformational exchange.[133] It was postulated that autoinhibition was achieved via the JMS "locking" the N and C terminal lobes of the KD in an unproductive conformation. Thus the additional

regulatory switch represented by the JMS could potentially also be a target for inhibitor design.

Dasatinib is a well-known kinase inhibitor that was approved for use in cases of imatinib resistance in chronic myeloid leukemia (CML).[154] Although dasatinib was primarily designed as a selective inhibitor of the Src and BCR-Abl kinases,[155] it was found to have a broader inhibition spectrum which includes members of the Eph family.[47, 141] However, no crystal structure of dasatinib in complex with a member of this family has thus far been elucidated. Since no structure of the wild type kinase domain was available, we wished to gain insight into the active form and to facilitate structure-based inhibitor design by determining the crystal structure of the EphA4 kinase domain in both the apo and dasatinib bound forms.

Material and methods

Plasmid construction

The murine EphA4 kinase domain (amino acids 606-846) was commercially synthesized. The gene was cloned into a pET28a vector (Invitrogen) generating an N-terminal 6His tag fused to the kinase domain with an intervening region containing an amino acid sequence efficiently hydrolyzed by the tobacco etch virus (TEV) protease. The purified protein contained a single non-coded glycine at the N terminus after cleavage.

Expression and Purification of EphA4 kinase domain

The EphaA4 kinase domain expression plasmid was transformed into *E. coli* BL21-CodonPlus (DE3)- RP competent cells, this specific strain contains extra copies of genes coding for rare codons R and P enabling efficient high expression of mammalian proteins that make use of these codons. Protein expression was induced overnight at 18 °C. A cell pellet corresponding to 1 liter of culture was resuspended in 30 ml of lysis buffer (50 mM Sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% Tween 20, 10% glycerol and 0.01% beta-mercapto-ethanol) supplemented with a protease inhibitor cocktail (complete EDTA-free) from Roche, lysed using a French press and centrifuged at 37,000 x g for 45 min at 4°C. The supernatant was applied to a 5 ml HisTrap HP column (GE Healthcare) with a peristaltic pump at 4°C for 1 hour (Binding buffer: 20

mM Tris-HCl pH 7.6, 500 mM NaCl and 5 mM imidazole). After extensive washing, the recombinant protein was eluted with binding buffer plus 500 mM imidazole. The 6His tag was removed by treatment with TEV protease (1:1 molar ratio) fused to a 6His tag expressed and purified according to published protocols[138] for 24 hr at 20°C in TEV activity buffer (25mM NaPi, 125 mM NaCl, pH 7.4, 5mM DTT). To remove TEV protease, the sample buffer was exchanged for binding buffer and then applied to a 5 ml HisTrap column. The unbound, cleaved EphA4 was incubated with 20 mM ATP and 20 mM $MgCl_2$ in order to fully activate it. After an hour at 4°C, the protein was diluted 10 fold in 10 mM HEPES pH 7.5, 10 mM NaCl and applied to a 1 ml HiTrap Q column (GE healthcare) that had been pre-equilibrated with the same buffer. The fractions containing the EphA4 kinase domain, which did not bind to the column, were then concentrated to 5 ml and applied to a Sephacryl S-200 HR column (HiPrep 16/60; GE Healthcare) equilibrated in 10 mM HEPES, 500 mM KCl and 1 mM DTT. This procedure yielded a protein that was homogenous as judged by Coomassie blue staining of SDS-PAGE gels and enzymatically active (not shown). A single purification from a liter of cell culture yielded between 3 and 6 mg of pure protein. Purified EphA4 kinase domain was concentrated to 11 mg/ml for crystallization using a Centricon-10 (Amicon).

Crystallization of the EphA4 kinase domain

Crystals were grown at 20°C using the sitting-drop vapor-diffusion method. The purified EphA4 kinase domain was mixed with an equal volume of a reservoir solution

containing 15-21% polyethylene glycol 10K, 100 mM ammonium acetate and 100 mM bistris pH 5.5. Rod-like crystals appeared after 1-3 days.

Co-crystallization of EphA4 kinase domain in complex with dasatinib

Dasatinib (*N*-(2-chloro-6-methylphenyl)-2-(6-(4-(2-hydroxyethyl)piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide) [156] was added to the purified protein at a final concentration of 1 mM in dimethyl sulfoxide (molar ratio dasatinib/protein of 3) and mixed with an equal volume of a reservoir solution containing 15-21% polyethylene glycol 10K, 100 mM ammonium acetate and 100 mM bistris pH 5.5. Crystals appeared after 1-3 days.

X-ray data collection, structure determination and refinement.

Crystals were flash frozen using the reservoir solution supplemented with 20% glycerol as a cryoprotectant. Diffraction data for the apo protein was collected at the Swiss Light Source (SLS, beamline X06SA) and at the European Synchrotron Radiation Facilities (ESRF, beamline ID23-2) for the inhibitor bound. Datasets at 1.5 Å for the apo-protein and 1.55 Å for the inhibitor complex were processed using XDS[157] and Pointless[158] from the CCP4i suite.[159] The apo-structure was solved by molecular replacement using the program AMoRe.[160] A mutant of EphA4 was used as a search model (PDB ID: 2HEL). From the AMoRe solution, a model was built and input into ARP/wARP[161] for automated tracing. The dasatinib-bound structure was solved by

molecular replacement (AMoRe) using the native protein as a search model. The structures were refined using rigid body refinement within the program REFMAC5.[162] This was followed by restrained refinement with PHENIX[163] interspersed with rounds of model building with COOT.[164] A summary of the data collection and refinements statistics is presented in Table 4.1.

Table 4.1: Data collection and refinement statistics

Data collection	Apo EphA4	Dasatinib-bound EphA4
PDB ID	2Y6M	2Y6O
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell constant α=β=γ=90°	A=32.41, b=91.64, c=98.35	a=32.41, b=91.64, c=98.35
Resolution (Å)	1.7	1.54
Total observations	87,554	173,478
Unique reflections	31,531	41,896
Completeness (%)	97.2 (97.3)[a]	95.4 (82.8)
R_{merge}[b](%)	4.6 (25.0)	6.9 (32.2)
Average I/σI	13.1 (4.1)	14.8 (4.5)
Refinement		
Rwork/Rfree (%)	18.61/21.95	18.01/20.05
Temperature factor (Å^2)	20.1	19.2
rmsd bond lengths (Å)	0.006	0.006
rmsd bond angles (°)	0.966	0.994
rmsd bond dihedrals (°)	11.864	16.84

[a]Values in parentheses are for the highest resolution bin.

[b]$R_{merge} = \Sigma_{hkl}\Sigma_j |I_{hkl,j} - \langle I_{hkl}\rangle| / \Sigma_{hkl}\Sigma_j I_{hkl,j}$, where I is the intensity for the j^{th} measurement of an equivalent reflection with the indices *h,k,l*.

Results & Discussion

Apo enzyme structure

The phosphorylated kinase domain of murine EphA4 (residues 606-846, referred to as EphA4 KD) was crystallized in the apo form and the structure was determined at 1.7 Å resolution. The crystallographic model contains one molecule in the asymmetric unit and was refined to an R_{work} =17.49% and R_{free}= 20.32%. The overall bilobed structure typical of kinases is well ordered except for three N-terminal residues and five C-terminal residues and residues 765-787 of the activation segment. The N-terminal lobe consists of the characteristic five-stranded β sheet and a single large α helix (αC) while the C-terminal lobe is larger and mainly α-helical, but includes a two stranded anti-parallel β-sheet as expected (Fig. 4.1). Protein kinases contain a number of conserved elements that contribute to the enzyme activity including: a hinge region connecting the two lobes, a glycine rich loop or p-loop, a catalytic loop and an activation loop, all of which are present in the EphA4 structures. The activation loop includes a consensus Asp-Phe-Gly (DFG) motif sequence that is critical for ATP binding[165] and is highly conserved among kinases.[166]

The present structure is characteristic of a phosphorylated, and therefore activated, form of a kinase domain.[167] First, the αC-helix is in the ATP proximal position allowing the formation of the conserved Lys^{653}-Glu^{670} salt bridge which coordinates the alpha and beta phosphate groups of ATP *via* polar contacts. Second, the activation loop is

disorganized and adopts the so-called "DFG-in" conformation[168] with the side-chain of Phe[765] pointing towards the interior of the molecule (Fig. 4.1B).

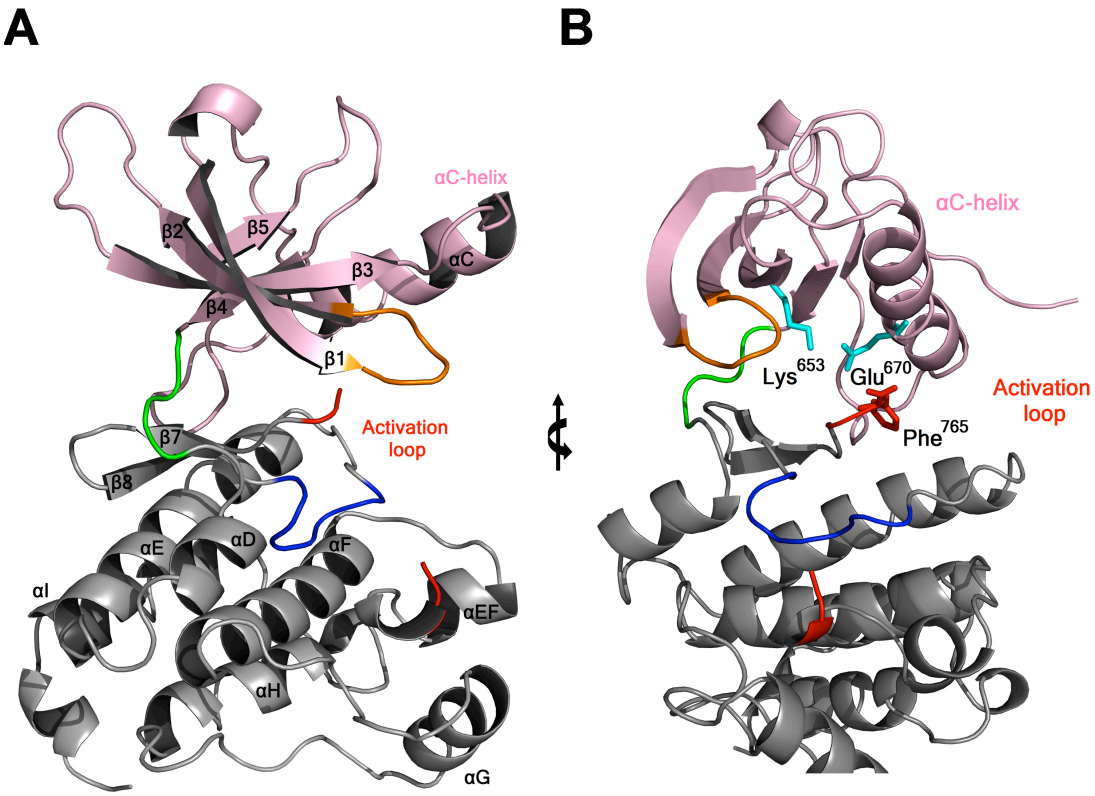


Figure 4.1: Overview of the crystal structure of the EphA4 kinase domain. (A) Ribbon representation of the apo-EphA4 kinase domain crystal structure. The N-terminal lobe, shown in pink, consists of a five-stranded anti-parallel β-sheet and one α-helix (αC). The C-terminal lobe, in grey, is mainly α-helical. The p-loop is colored orange, the hinge region in green, the catalytic loop in blue and the activation loop in red. (B) Same as in A with a 90° rotation about the z axis. Residues important for kinase activation (Lys[653], Glu[670] and Phe[765]) are shown in stick representation. All molecular graphics were generated with PyMOL.[169]

Comparison of Eph receptor crystal structures.

An EphA4 protein consisting of the kinase domain and the adjacent juxtamembrane segment (JMS), which comprises the mutations Y596/602F in the JMS and Y742A in the KD, has been previously crystallized (PDB ID: 2HEL).[133] We refer to this structure as EphA4 JMS-KD. Although the resolution of this structure is 2.35 Å, the JMS is almost completely disordered. Overall, the structures of the two proteins are very similar with an r.m.s. deviation of 0.633 Å for the superimposition of 208 Cα atoms (Fig. 4.2A and Table 4.2). The C terminal lobe is nearly identical in the two structures with the exception that five less amino acid residues of the activation loop of EphA4 KD could be modeled. As the catalytic and activation loops are spatially close to each other in the EphA4 KD structure, the presence of the bulky side-chain of Tyr^{742} in the catalytic loop leads to reduced interaction with the activation loop and its subsequent disorganization. The Y742A mutation in EphA4 JMS-KD removes the tyrosine side-chain allowing repacking and stabilization of the activation loop. In contrast, the N-terminal lobe shows greater differences between the KD and JMS-KD structures. The largest difference observed is in the location of the αC-helix where a 4.5 Å movement of the Cα of Asp^{661-} upward and outward in the KD structure causes the displacement of the entire alpha helix. The αC-helix contains a kink that leads to a bend of the long axis by 14° at about the midpoint of the helix (Fig. 4.2A). The kink is similar in both structures, however the relative position of the alpha helix is different, which in turn influences the position of the conserved Lys-Glu salt bridge. The salt bridge is also influenced by the position of the p-loop which is shifted upwards in the KD structure compared to the JMS-KD structure. Thus, although

the two crystals structures share an almost identical amino acid sequence, there are significant structural differences.

To try and determine whether the structural differences between EphA4 KD and JMS-KD are related to the presence of the JMS, we compared this pair to the structures of another KD JMS-KD pair from the Eph family. Crystal structures of the closely related kinase EphB2 have been solved both in the active state (EphB2 KD, PDB ID: 2HEN) which comprise a D754A mutation in the activation loop)[133] and the auto-inhibited state (EphB2 JMS-KD, PDB ID: 1JPA).[132] In the EphB2 JMS-KD structure, a Y604/610A mutation in the JMS has been made to prevent activation, which results in an enzyme with constitutively low activity.[131] Overall, the structures of the 2 pairs of kinases are similar except for the αC-helix and the p-loop which present slight changes (Fig. 4.2B and Table 4.2). The αC-helix exhibits a similar kink in all the structures, the cause of which was previously attributed to the JMS.[132] However, a subsequent crystal structure of EphA2 was reported which also exhibited a bent conformation of the αC-helix despite the absence of a JMS (Fig. 4.2C).[170] The situation is similar with the present crystal structure of EphA4 KD, i.e. a kink in the αC-helix even though the JMS is absent. This observation confirms the previous suggestion[133] that the presence of the JMS in the EphA4 JMS-KD structure is not responsible for the kink in the αC-helix. However, the relative position of the αC-helix is different in all the structures (Fig. 4.2C). In the present structure of active EphA4, the position of the αC-helix is closer to the auto-inhibited JMS-KD of EphB2, indicating that the details of the position of the αC-helix do not necessarily correlate with the state of activity of the kinase.

The variability of the p-loop in all 4 structures shown in Fig. 4.2C, which influences the position of the conserved Lys-Glu salt bridge, is even greater than the αC-helix. In the kinase domain of EphA4, the position of the salt bridge is shifted, creating a larger hydrophobic pocket than in the other Eph receptor kinases, including specifically the EphA4 JMS-KD (Fig. 4.2D). Other Eph receptor kinase domains that have been crystallized without the JMS or mutations lack this hydrophobic pocket due to the position of the salt bridge (Fig. 4.3). In addition, it seems likely that this pocket is present in the active, full-length kinase considering inhibition of EphA4 by dasatinib has been demonstrated in the cell[141] and dasatinib occupies this back pocket in our crystal structure (see below). This newly discovered back-pocket should allow for both enhanced specificity and potency for EphA4 inhibitors.

Table 4.2: Comparison of the crystals structures of EphA4 and EphB2 receptors

PDB ID	Protein	Structure	r.m.s.d. versus EphA4 kinase domain		
			Overall	N lobe (aa 609-704)[a]	C lobe (aa 705-891)[a]
2HEL	EphA4	JMS-KD	0.633 Å^2 (208 Cα)	0.581 Å^2 (76 Cα)	0.267 Å^2 (139 Cα)
2HEN	EphB2-molecule A	KD	0.698 Å^2 (214 Cα)	0.577 Å^2 (80 Cα)	0.298 Å^2 (143 Cα)
	EphB2-molecule B	KD	0.818 Å^2 (216 Cα)	0.582 Å^2 (81 Cα)	0.282 Å^2 (147 Cα)
	EphB2-molecule C	KD	0.734 Å^2 (217 Cα)	0.534 Å^2 (78 Cα)	0.285 Å^2 (141 Cα)
	EphB2-molecule D	KD	0.838 Å^2 (217 Cα)	0.582 Å^2 (80 Cα)	0.282 Å^2 (145 Cα)
1JPA	EphB2-molecule A	JMS-KD	0.472 Å^2 (223 Cα)	0.617 Å^2 (83 Cα)	0.278 Å^2 (153 Cα)
	EphB2-molecule B	JMS-KD	0.630 Å^2 (221 Cα)	0.637 Å^2 (80 Cα)	0.312 Å^2 (143 Cα)

[a] The number of Cα atoms used for superposition is given in parenthesis.

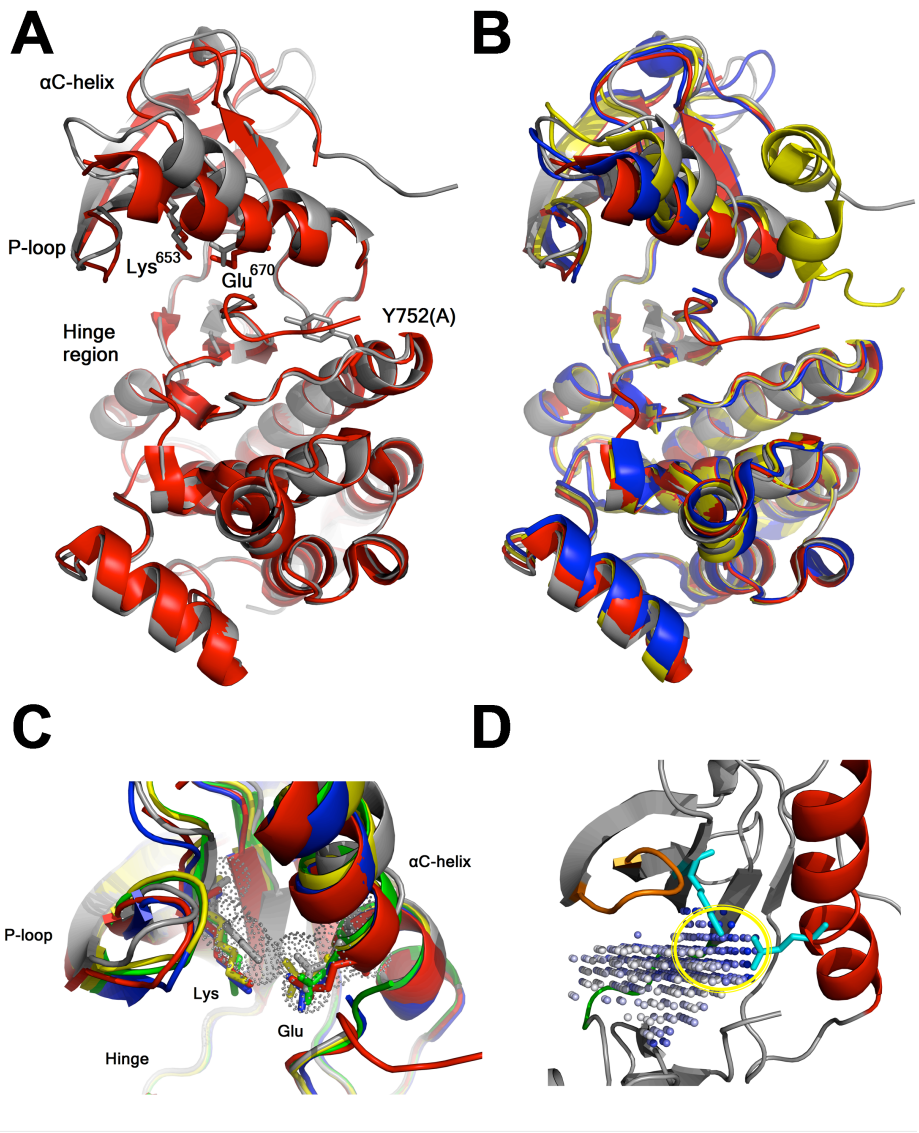


Figure 4.2: Comparison of the EphA4 crystal structure with the kinase domain of other Eph receptor. (A) Superimposition of EphA4 structures: EphA4 KD (grey), EphA4 JMS-KD (2HEL, red); Shown in bond representation: the residues contributing to the conserved salt bridge Lys-Glu and the Tyr/Ala residue in the activation loop (B) Superimposition of EphA4 and EphB2 structures: EphA4 KD (grey), EphA4 JMS-KD (2HEL, red), EphB2 KD (2HEN, blue), EphB2 JMS-KD (1JPA, yellow). (C) Superimposition of EphA4, EphB2 and EphA2: closer view of the N-lobe, the residues contributing to the conserved salt bridge Lys-Glu are shown in bond representation while the dots represent the van der Waal's surface of the salt bridge. EphA4 KD (grey), EphA4 JMS-KD (2HEL, red), EphB2 KD (2HEN, blue), EphB2 JMS-KD (1JPA, yellow), EphA2 KD (1MQB, green). (D) Close up view of the ATP binding pocket of EphA4. The dots show the ATP binding pocket and the back hydrophobic pocket is circled in yellow. The orange, green, cyan and red elements represent respectively the p-loop, hinge region, the salt bridge Lys^{653}-Glu^{670} and the αC helix.

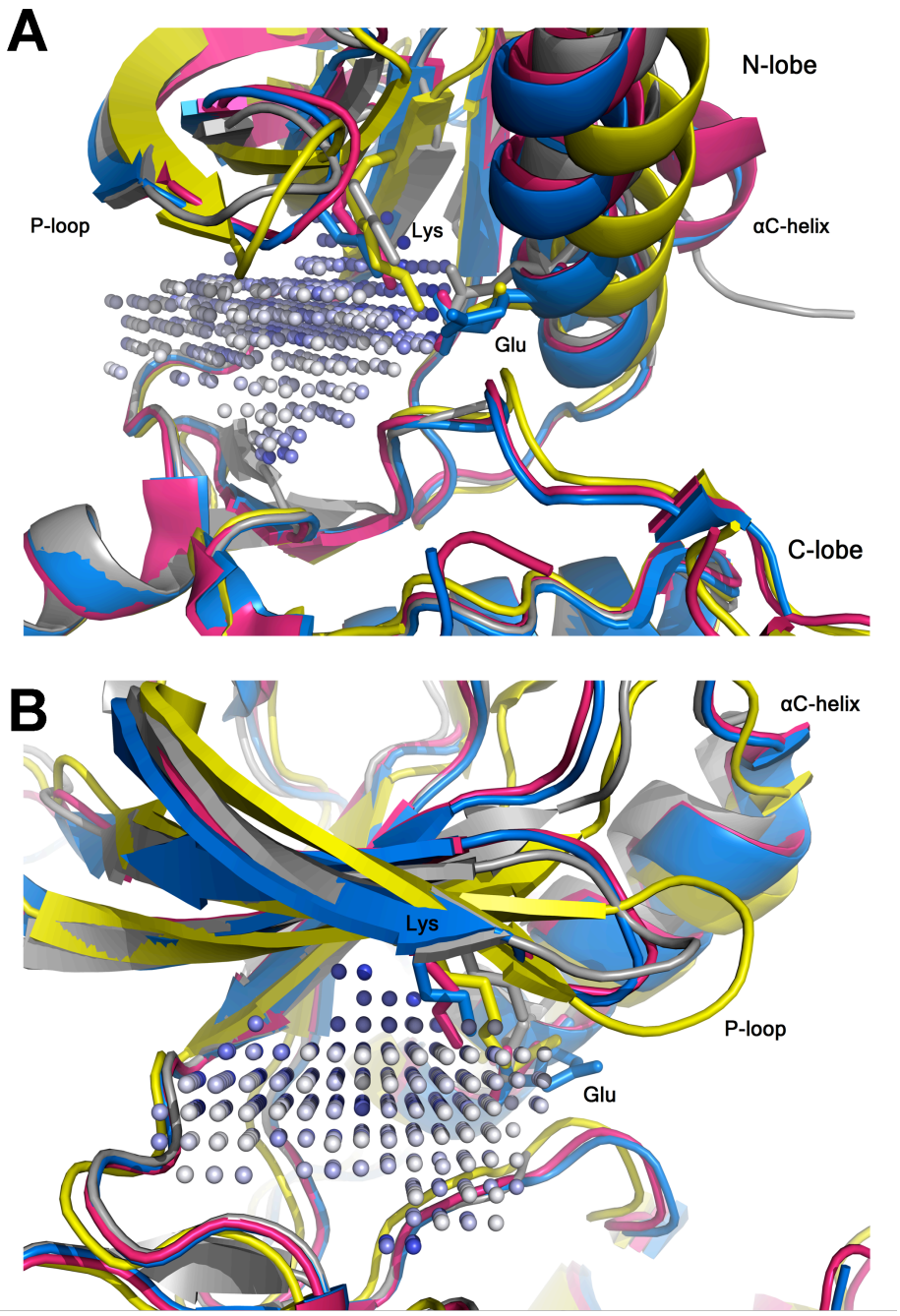


Figure 4.3: Comparison of the EphA4 crystal structure with the kinase domain of other Eph receptors from the protein databank containing no mutations. (A) Superimposition of structures: EphA4 KD (grey), EphA3 KD (2QOB, blue) EphA5 (2R2P, magenta) EphA8 (3KUL, yellow); Shown in bond representation: the residues contributing to the conserved salt bridge Lys-Glu while the dots represent the ATP binding pocket of EphA4. (B) Same representation but counter clockwise rotated about the z axis by an angle of 20°.

Structure of EphA4-dasatinib complex

Dasatinib is an ATP competitive inhibitor that targets multiple kinases. Amongst the numerous targets of dasatinib, it has been shown to inhibit EphA4.[47, 141] Using a commercially available kinase assay (ADP Hunter Plus, DiscoveRx), the IC_{50} of dasatinib for the kinase domain of EphA4 was 25 nM (cf. Chapter 5, Fig. 5.1). However, no structural information regarding the binding of dasatinib to EphA4 is available. In order to address this issue, dasatinib was co-crystallized with EphA4 KD under conditions identical to the apo protein. The structure of the complex was solved to 1.55 Å resolution and refined to an R_{work} = 18.01 % and R_{free}= 20.05% (Fig. 4.4 and Table 2). Structural alignment of the apo and the dasatinib bound forms of EphA4 KD gives an r.m.s deviation of 0.150 Å over 220 Cα atoms. Although the 3D structures are very similar, the B factors of the apo and inhibitor bound proteins are subtly different (Fig. 4.5). B-factors are a measure of the order in the unit cell. Careful inspection reveals that the backbone atoms have on average slightly lower B-factors in the inhibitor bound complex while the side-chains have somewhat larger B-factors. Dynamic behavior in solution may give rise to increased B-factors. Interestingly, a similar pattern of dynamic changes on ligand binding (decreased dynamic behavior of the backbone and increased dynamics of the sidechains) has been observed in NMR.[171] The pattern of changes in the B-factors suggests a thermodynamic explanation for binding i.e. that reduced entropy of the backbone in the dasatinib bound form is compensated by increased entropy in the sidechains. As a result, the net entropic change of ligand binding is minimal. The reduced B-factors for backbone atoms extends slightly further towards the N-terminus in the

dasatinib bound protein allowing two additional amino acids to be modeled. As with the apo protein, the αC-helix is in the ATP proximal position allowing the formation of the conserved Lys^{653}-Glu^{670} salt bridge indicative of the kinase active state[167] while the activation loop is also in the "DFG-in" conformation.[168]

Dasatinib makes several hydrogen bonds to the hinge and occupies the hydrophobic pocket behind the Thr^{699} residue which is commonly referred to as the 'gatekeeper' residue.[172] The side-chain of the gatekeeper residue sterically controls inhibitor binding to the hydrophobic pocket.[173] There are three main hydrogen bonds between dasatinib and the kinase: one between the nitrogen of the amino-thiazole ring and the amide nitrogen of Met^{702}, another between the amino group of the amino-thiazole ring with the carbonyl oxygen of Met^{702} and a third one between the amide nitrogen of dasatinib and the side-chain hydroxyl oxygen of Thr^{699}. The 2-chloro-6-methyl phenyl group of dasatinib occupies the hydrophobic pocket and makes extensive van der Waal's (vdW) contacts with Lys^{653}, Glu^{670}, Met^{674}, Ile^{683}, Ile^{697} and Thr^{699}. The amino-thiazole ring has vdW's contacts with Leu^{753} and Ala^{651}, in addition to the previously described hydrogen bond. The pyrimidine ring extends out of the binding pocket and makes vdW contact with Ile^{627}. The hydroxyethyl-piperazine group is largely solvent exposed.

In an analysis of the Eph receptor family,[47] it was observed that, except for EphA6 and EphA7, dasatinib inhibited all of the Eph receptors. Interestingly, the gatekeeper residue in EphA6 is Val and Ile in EphA7, while in all other members of the family it is Thr. The bulky gatekeeper residues in EphA6 and EphA7 likely obstruct access to the hydrophobic pocket. The pattern of inhibition is not surprising since dasatinib was designed to access the hydrophobic pocket created by the mutation of the gatekeeper

residue from Ile to Thr in Abl kinase that lead to resistance to imatinib. It seems likely that the mode of binding of dasatinib observed here is similar for the other Eph receptors that are inhibited by this compound.

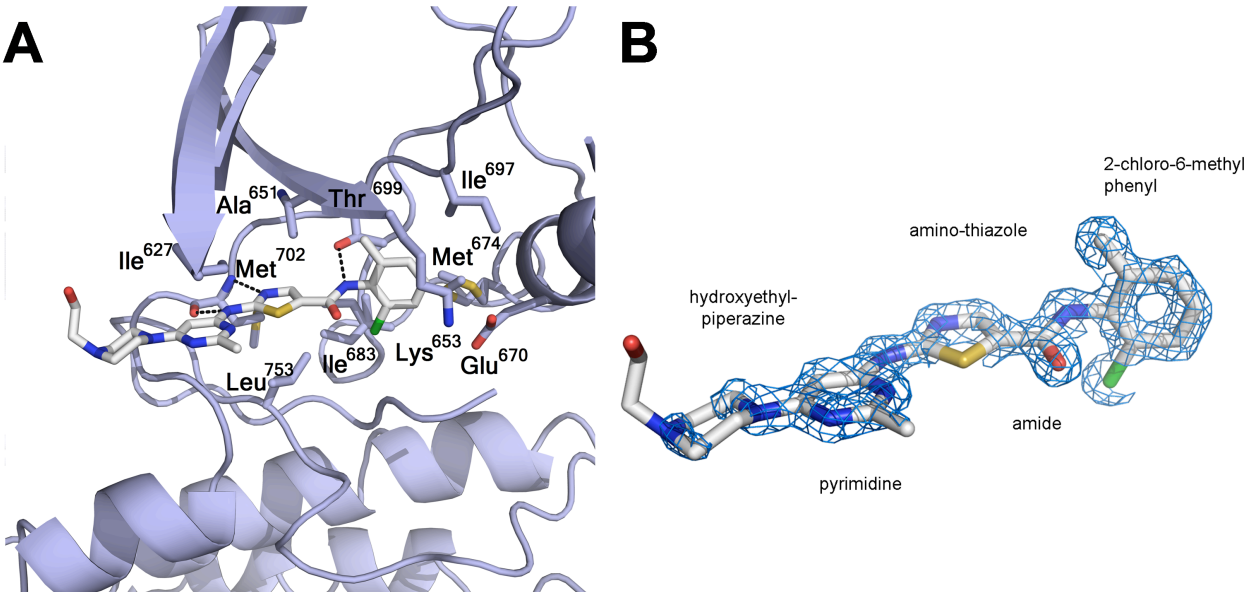


Figure 4.4. Binding mode of dasatinib to EphA4. (A) Overview of the binding interaction. Dasatinib and residues that interact with it are shown in bond view, hydrogen bounds are indicated by black dashed lines. (B) View of dasatinib overlaid with the simulated annealing omit-map F_0-F_c electron density contoured at 2 σ.

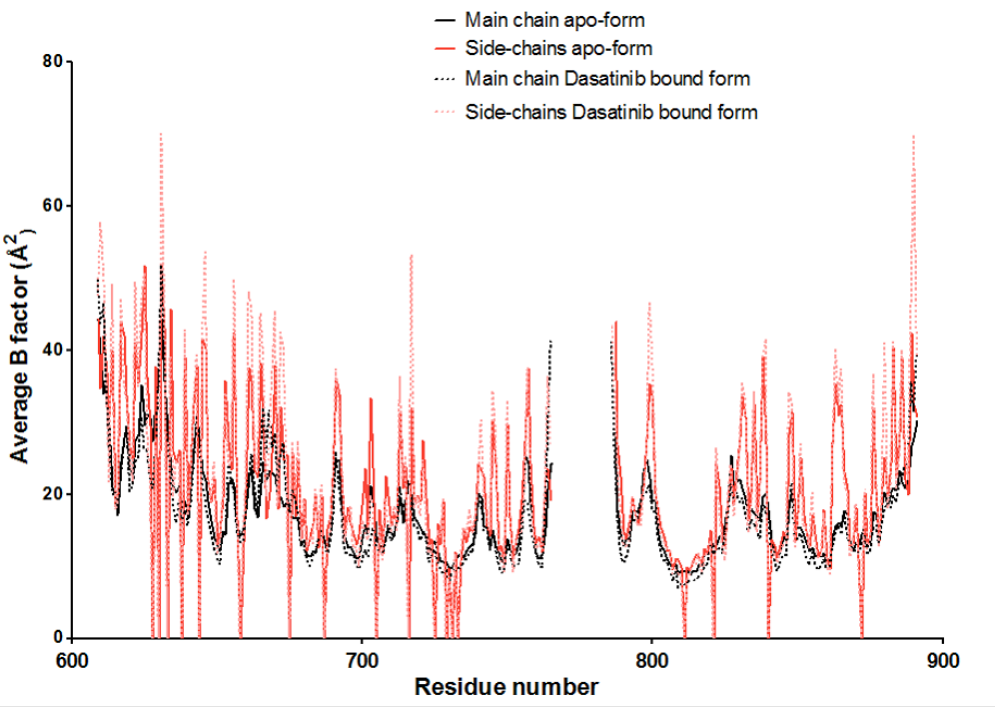


Figure 4.5. Average B factor per residue number. In black the values for main chain, and in red values for the side-chains. Solid lines are values for the apo-form and the dashed lines are values for the dasatinib bound form.

Structural comparison with other kinases bound to dasatinib.

Since a number of other dasatinib bound structures have been elucidated, it is useful to compare the present structure to these. We have selected four representative structures, two from the Src family (Src[174], and Lyn[175]), the Ableson kinase (Abl)[176] and Bruton's tyrosine kinase (BTK)[177] for comparison with our dasatinib-bound EphA4 structure. In all five structures, the 2-chloro-6-methyl phenyl group of dasatinib occupies a hydrophobic pocket adjacent to the ATP-binding site behind the gatekeeper residue, which in all cases is a Thr (Fig. 4.6A). All 5 structures exhibit an αC-helix in the ATP-proximal position,

the salt bridge between Lys and Glu and an activation loop with the "DFG-in" conformation. The kink in the αC-helix is present in all the structures and the position of the helix is slightly different in each (Fig. 4.6B). However, the salt bridge is in the same position in all the structures. Although the binding mode of dasatinib is globally similar in all the structures, two different binding orientations are observed for the 2-chloro-6-methyl phenyl group of dasatinib. The chloride is pointing toward the C-terminal lobe in the Src, Lyn and EphA4 complexes (Fig. 4.6C), while the chloro substituent is pointing in the opposite direction in the Abl and BTK complexes (Fig. 4.6D). However, it is not clear why the different binding orientations are observed, it may be related to the resolution. In the present, relatively high resolution EphA4 structure, the orientation of the chloro substituent pointing toward the C-terminal lobe fit the electron density better.

In addition to the hydrophobic pocket, differences are observed in the p-loop in both Abl and BTK structures compared to EphA4. There is a 7.7Å movement of the Cα of Abl Gly^{250} towards the ATP binding site in comparison to the corresponding EphA4 Gly^{630} (Fig. 4.6D). The p-loop of BTK shows a similar movement to Abl, although to a slightly lesser extent, with a 3.65Å movement of the Cα of BTK Gly^{411} towards the ATP binding site in comparison to the corresponding Gly^{630} of EphA4. Moreover, Abl Tyr^{253} makes several vdW's contacts with the amino thiazole of dasatinib that are absent in the other complexes. The p-loop in BTK and Abl packs closer to the core of the kinase as evidenced by the shorter distance between the methyl substituent of the pyrimidine ring and the p-loop of BTK (8.69 Å to Gly^{411}) and Abl (7.12 Å to Gly^{250}) in comparison to EphA4 (9.85 Å to Gly^{630}), thereby partially shielding dasatinib from solvent. The position of the p-loop may relate to the differences in affinity for dasatinib noted in Fig. 4.6A.

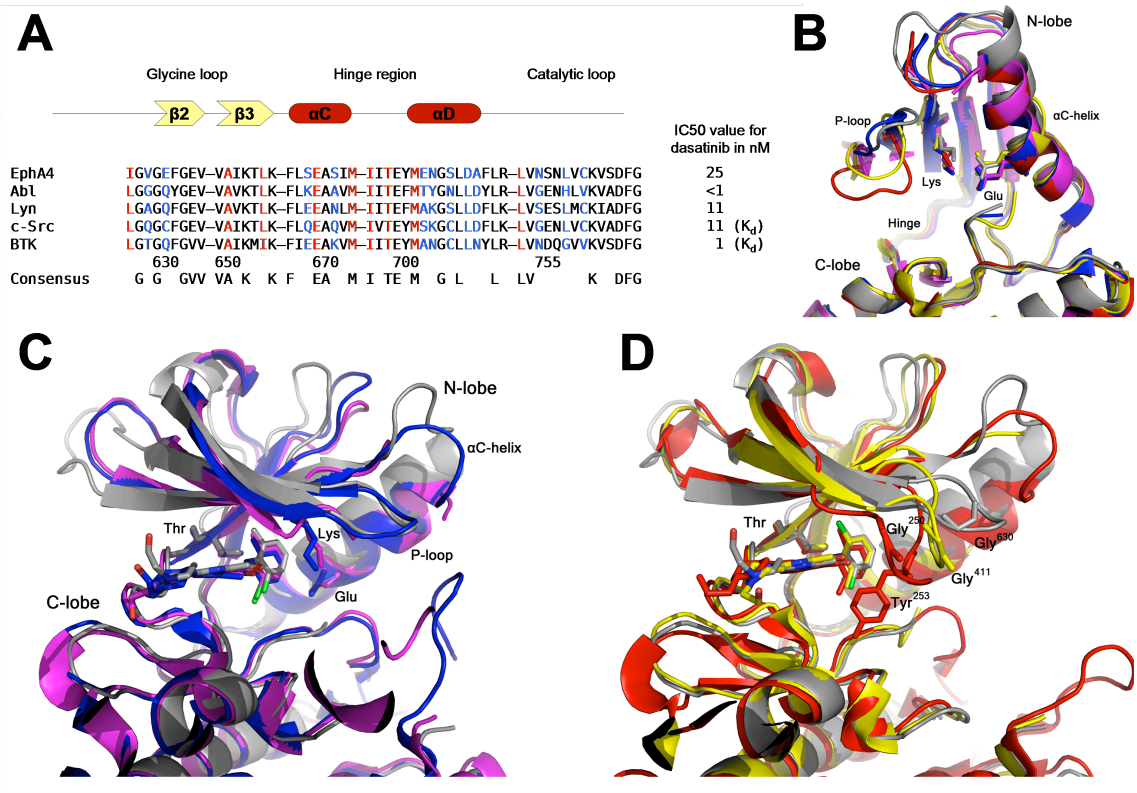


Figure 4.6. (A) Sequence alignment of a portion of the EphA4 kinase domain with other RTKs that bind dasatinib: Lyn, BTR, c-Src and Abl. Variable residues are indicated in blue; and the residues that interact with dasatinib are shown in red. The potency of dasatinib inhibition of each kinase is indicated on the right. (B) Superimposition of EphA4 (2Y6O, grey), Lyn (2ZVA, blue), c-Src (3G5D, magenta), BTK (3K54, yellow) and Abl (2GQG, red). The residues contributing to the conserved salt bridge are shown in bond representation. (C) Superimposition of EphA4 (2Y6O, grey), Lyn (2ZVA, blue) and c-Src (3G5D, magenta). The residues contributing to the conserved salt bridge and the gatekeeper residue (Thr) are shown in bond representation. (D) Superimposition of EphA4 (2Y6O, grey), BTK (3K54, yellow) and Abl (2GQG, red), in addition to the residues contributing to the salt bridge and the gatekeeper residue (Thr), selected residues in the p-loop are shown in bond representation (Gly^{250} and Tyr^{253} for Abl, Gly^{411} for BTK and Gly^{630} for EphA4).

Conclusion

We have elucidated the structure of the native EphA4 kinase domain and the first structure of this domain in complex with an inhibitor. The structure of the native kinase domain suggests that the JMS in EphA4 is not responsible for the change in the relative position of the αC-helix as was previously suggested [133]. In addition, the structure reveals a hydrophobic back-pocket in the ATP-binding site of EphA4 which was unknown before. The crystal structure of EphA4 in complex with dasatinib revealed a binding mode closely related to the one exhibited by Src family members, c-Src and Lyn.

Access to the hydrophobic pocket is governed by the gatekeeper residue while the size of the pocket appears to be governed by the position of the Lys-Glu salt bridge. The gatekeeper residue is conserved among 12 of the 14 human Eph receptors while, in our structure, the position of the conserved salt bridge is shifted, offering a larger hydrophobic pocket than in the other Eph receptors. This unique combination of gatekeeper and position of the salt bridge suggests a manner to design new, EphA4 specific inhibitors. Further mutational experiments targeting the ATP binding site and the hydrophobic pocket in particular could shed light on the potential to develop EphA4 specific compounds.

PDB accession numbers

Atomic coordinate and structure factors of apo- EphA4 and dasatinib-EphA4 were deposited in the Protein Data Bank under accession numbers 2Y6M and 2Y6O respectively.

Chapter 5

Discovery and structure of an *in silico* fragment inhibitor

The *in silico* identification, optimization and crystallographic characterization of a 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine scaffold as an inhibitor for the EphA4 receptor tyrosine kinase is described. A database containing commercially available compounds was subjected to an *in silico* screening procedure which was focused on finding novel, EphA4 hinge binding fragments. This resulted in the identification of 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine derivatives as EphA4 inhibitors. Hit exploration yielded a compound with 2 μM (IC_{50}) affinity for the EphA4 receptor tyrosine kinase domain. Soaking experiments into a crystal of the EphA4 kinase domain yielded a 2.11Å X-ray structure of the EphA4 – inhibitor complex, which confirmed the binding mode of the scaffold as proposed by the initial *in silico* work.

This chapter is based on the following publication in which my contribution to the work involved crystallography and structure analysis.

O. P. van Linden, C. Farenc, W. H. Zoutman, L. Hameetman, M. Wijtmans, R. Leurs, C. P. Tensen, G. Siegal and I. J. de Esch, Fragment based lead discovery of small molecule inhibitors for the EPHA4 receptor tyrosine kinase, Eur J Med Chem, 2012, 47, 493-500.

Introduction

The erythropoietin-producing human hepatocellular (Eph) carcinoma receptor tyrosine kinase family is the largest sub-family of receptor tyrosine kinases. These membrane spanning receptors and their ligands (ephrins) play critical roles in the developmental stage of vertebrates and are involved in processes such as axon guidance and angiogenesis.[51] There are nine EphA receptors and five EphB receptors in the human genome. This classification is based on early insights with respect to the activating ligands. It was believed that the EphA receptors promiscuously interact with glycosylphosphatidylinositol-anchored ephrin-A ligands, while the EphB receptors promiscuously interact with the transmembrane ephrin-B ligands. More recently it has been shown that the EphA4 receptor and the EphB2 receptor are exceptional as they are capable of binding both A and B types of ephrins.[178, 179] It has been demonstrated that the Eph receptors and ephrins are overexpressed in various tumor types.[51] For example, the EphA4 receptor, normally only expressed in the developing brain, was found to be highly overexpressed in tumor cells of cutaneous T-cell lymphomas (CTCL),[180] prostate cancer and pancreatic cancer.[181, 182] The potential oncogenic properties of EphA4 became evident from studies on prostate and pancreatic cancer cell lines, where down regulation of EphA4 with siRNA reduced cell viability.[183] These results suggest that EphA4 might be a promising target for pharmacologic intervention. Furthermore, it has been suggested that EphA4 inhibition might be beneficial for the treatment of spinal cord injuries.[184, 185] These data validate the concept of developing EphA4 inhibitors as potential drugs and selective EphA4 blockers are therefore desired. Although high affinity kinase inhibitors show cross

activity towards EphA4[186, 187] and small molecules can antagonize ephrin binding and EphA4 activation,[188] none of these compounds has any specificity for EphA4. Therefore, our aim was to identify novel scaffolds capable of specifically inhibiting the EphA4 kinase activity. Towards this end, we designed a detailed pharmacophore screening model, which targets the well described hinge region and gatekeeper[189] residue in the ATP binding site. By using an *in silico* screening cascade, a database containing a large number of commercially available compounds was reduced to a small set of hit scaffolds, which were tested for their EphA4 inhibitory properties.

Kinase Inhibition Assay

Kinase activity was measured using the ADP Hunter Plus assay kit according to the manufacturers (DiscoveRx, Birmingham, UK) recommendations. In this assay ATP turnover, and the subsequent generation of ADP, is measured in a two-step process. Reaction mixtures (20 μl) were prepared in black polypropylene 384 well flat bottom plates (Greiner, Stonehouse, UK) and contained 100 nM His-tagged EphA4-KD, 33 μM ATP (determined as Km for this kinase preparation), 200 μg/mL E4:Y1 (Sigma, St. Louis, MO) as substrate and increasing concentrations of inhibitors. Assay buffer (pH 7.4) contains 20 mM NaCl, 15 mM HEPES, 20 mM $MgCl_2$, 1 mM EGTA and 0.1% Tween20. After 75 min at 30ºC, reagents A and B from the kit were added and the fluorescence intensity was measured with a Tecan infinite F200 Reader (Tecan, Grödig, Austria) after 60 min incubation at 30 ºC, with an Ex of 535 nm and Em of 590 nm and 25 reads per well. Standard curves, ranging from 30 nM to 75 μM ADP (dynamic range of the kit), were used as an internal assay control. Background was measured in reactions containing no substrate and subtracted from those containing E4:Y1. Dasatinib was used as a positive control. For confirmed hits, the effects of the inhibitors on the enzymes in the ADP detection step were determined in assay buffer and the reagents A and B only.

Crystallography

The crystal structure of apo EphA4 kinase domain and the kinase domain bound to dasatinib have been elucidated.[190] Briefly, purified apo kinase domain was used to grow crystals at 20°C using the sitting-drop vapour-diffusion method. The purified EphA4 kinase domain was mixed with an equal volume of a reservoir solution containing 15-21% polyethylene glycol 10K, 100 mM ammonium acetate and 100 mM bistris pH 5.5. Rod-like crystals appeared after 1-3 days. Crystals of apo protein were soaked in drops of compound **73** at 30 mM. After 6 hours incubation at 20°C, crystals were flash-frozen using 20% glycerol plus the reservoir solution. Diffraction data was collected at 100 K at the ESRF. Datasets at 2.12 Å resolution were merged and processed using XDS[142] and Pointless[143] from the CCP4[144] suite and the structure was solved by molecular replacement using the program AMoRe.[145] The apo EphA4 kinase domain was used as a search model. The structure was subsequently refined using the program PHENIX[146] and the program COOT[147] was used for model building. An overview of the quality of the structure is presented in Table 5.1.

Table 5.1: Data collection and refinement statistics.

Data collection	EphA4 in complex with compound **27**
PDB ID	2XYU
Space group	$P2_12_12_1$
Cell constant α=β=γ=90°	a=32.54, b=91.54, c=97.95
Resolution (Å)	2.12
Total observations	17378
Unique reflections	16232
Completeness (%)	99.5 (98.2)[a]
R_{merge}[b](%)	13.8 (69)
Average I/σI	8.5 (1.9)
Refinement	
Rwork/Rfree (%)	18.94/22.98
Temperature factor (Å^2) (Å)	20.96
rmsd bond lengths (Å)	0.009
rmsd bond angles (°)	1.024
rmsd bond dihedrals (°)	12.221

[a]Values in parentheses are for the highest resolution bin.

[b]$R_{merge} = \Sigma_{hkl}\Sigma_j |I_{hkl,j} - \langle I_{hkl}\rangle| / \Sigma_{hkl}\Sigma_j I_{hkl,j}$, where I is the intensity for the j^{th} measurement of an equivalent reflection with the indices *h,k,l*.

Results

The *in silico* screening of the EphA4 kinase domain, design and synthesis of new ligand as well as docking studies were performed by the Ph.D. candidate Oscar van Linden in the group of Dr. de Esch.

The screening model describes the pharmacophore features needed for interactions between the EphA4 kinase domain (more specifically, the ATP binding site) and putative ligands. The apo structure of EphA4 (PDB ID: 2HEL)[191] was one of the templates used to construct the model. To accurately define ligand-receptor interacting features, the apo-EphA4 structure was superimposed on the crystal structures of two related kinases.[192] These crystal structures contain ligands which are known to have affinity for EphA4, namely AMPPNP and dasatinib (Fig. 5.1). AMPPNP, the structural analog of the endogenous substrate ATP, is expected to bind to the kinase domain in a similar fashion as ATP. Meanwhile, quantitative analysis of a variety of inhibitors using a panel of kinases revealed that the drug dasatinib has high affinity for the EphA4 receptor.[187] The structure of EphA3 co-crystallized with AMPPNP and the structure of ABL1 co-crystallized with dasatinib were superimposed on the apo-structure of EphA4. The binding mode of AMPPNP was derived from the EphA3-AMPPNP complex **I** (PDB ID:2QO9, Fig. 5.2A).[193] The AMPPNP adenine ring interacts with the kinase hinge region by forming two hydrogen bonds with the backbone atoms of residues Met^{702} and Glu^{700}. A third, water mediated, hydrogen bond to the adenine ring is formed by Thr^{699}, the kinase gatekeeper residue. The ribose hydroxyl groups and phosphate oxygen atoms make additional hydrogen bonds to the protein, which will not be considered here.

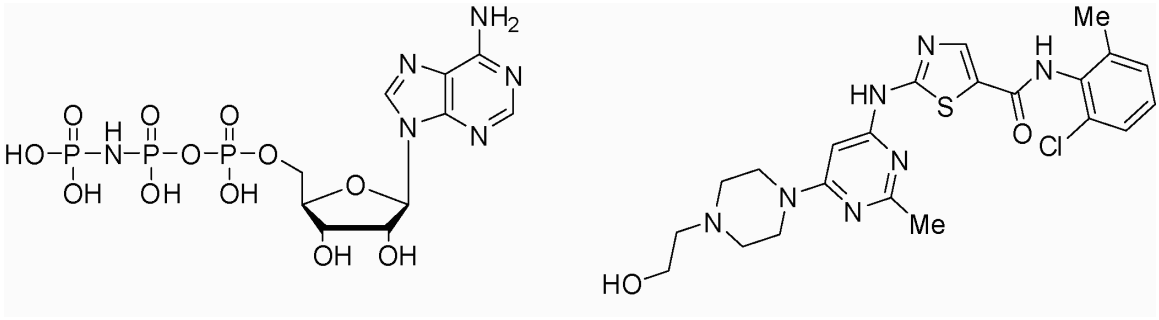


Figure 5.1. Structures of AMPPNP (left) and dasatinib (right).

From inspection of the Abl1-dasatinib complex **II** (PDB ID: 2GQG, Fig. 5.2B),[194] it is evident that the thiazole-2-amine moiety of dasatinib makes two hydrogen bonds with the kinase hinge region by interacting with the backbone atoms of Met[318]. An additional hydrogen bond to the hydroxyl group of the gatekeeper residue Thr[315] is formed by the amide nitrogen. The orthogonally oriented 2-chloro-6-methylphenyl ring probes the ABL1 kinase hydrophobic back-pocket, a pocket that is often used in kinase inhibitor development to improve potency and increase selectivity over other kinases.[189] The apo crystal structure of EphA4 was aligned with complexes **I** and **II**. The detailed side and top views of the superposed complexes (Fig. 5.2C and 5.2D) illustrate the binding modes of both ligands.

The MOE-2007 Pharmacophore Query Editor was used for the generation of the pharmacophore query based on the superposed complexes. Combining the pharmacophore features of AMPPNP and dasatinib resulted in the identification of the well-known donor–acceptor–donor motif,[195] visualized by the pharmacophore features F1, F3 and F5 (Fig. 5.2E). On the protein side, the corresponding features to F3 and F5

are shown (F2 and F6 respectively). These corresponding features in the screening model assure the correct directionality of a potential hydrogen bond donor / acceptor. The amide nitrogen atom of dasatinib forms a hydrogen bond with the gatekeeper residue Thr^{699}, by donating a hydrogen to the hydroxyl oxygen atom of this threonine residue. This is indicated by pharmacophore feature F7. The threonine gatekeeper, however, is able to interact with ligands by either accepting or donating a hydrogen atom for hydrogen bonding. Pharmacophore feature F7 is therefore annotated as donor or acceptor and F8 describes the corresponding feature to F7. Only ~20% of the eukaryotic protein kinases have a threonine residue as gatekeeper[196] and specifically addressing interactions with this residue might therefore be beneficial in terms of selectivity.

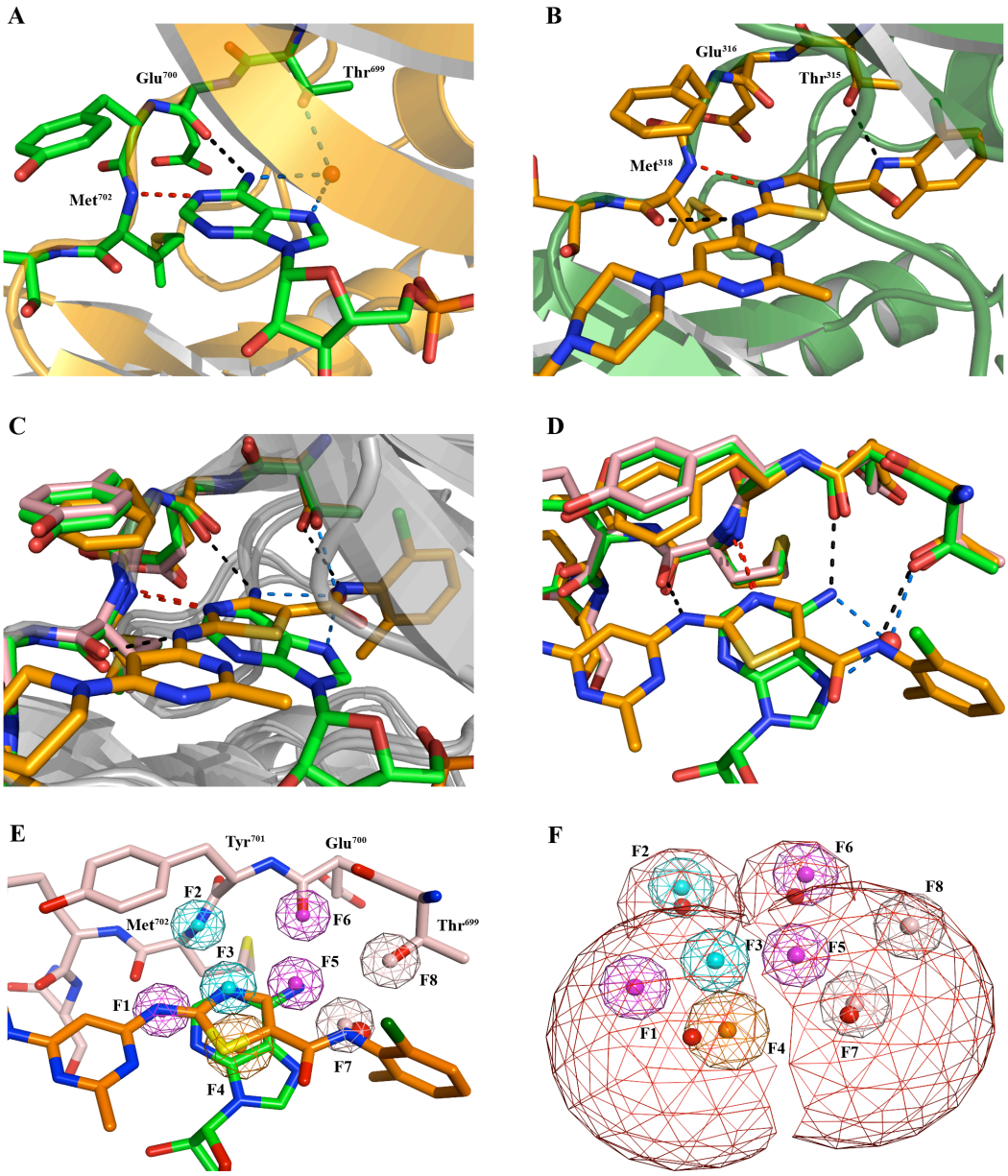


Figure 5.2: **A**, binding mode of AMPPNP in EphA3 (complex **I**, PDB ID: 2QO9). **B**, binding mode of dasatinib in Abl1 (complex **II**, PDB ID: 2GQG). **C**, superposition of complexes **I** (green carbon atoms), **II** (orange carbon atoms) and the apo structure of EphA4 (light pink carbon atoms, PDB ID: 2HEL). **D**, detailed top view of superposed complexes **I** (green carbon atoms), **II** (orange carbon atoms) and apo-EphA4 (light pink carbon atoms). **E**, annotation of pharmacophore features; F1, F5 = H-bond donor, F6 = H-bond donor projection, F3 = H-bond acceptor, F2 = H-bond acceptor projection F4 = aromatic center, F7 = H-bond donor or acceptor, F8 = H-bond donor or acceptor projection. **F**, pharmacophore screening model, constrained by the *exterior volume*. The excluded volume on the amino acid residues that make up the shape of the binding site is not shown for clarity of presentation. Annotations of pharmacophore features are similar to that in Figure 5.4.

The cores of the adenine and thiazole rings make up the aromatic feature F4. To constrain the size of potential hit compounds, an exterior volume filter was created (Fig. 5.2F) in the center of the ATP binding site. Around the carbon atom 5 of the dasatinib thiazole ring, a box of 11x10x7 (LxWxH) Angstrom was drawn in which the exterior volume was placed. The shape of the binding site was defined by the addition of an excluded volume which matched the van der Waals radii of the atoms from the residues that make up the binding site. These efforts led to a pharmacophore model that consists of eight features, which was used for the *in silico* screening of a database containing commercially available compounds

In silico screening cascade

The MOE-2007 Vendor Compound 3D Collection containing 2.7 million commercially available compounds, which were conformationally pre-sampled into 171.8 million conformations, was subjected to an *in silico* screening cascade (Fig. 5.3). A pharmacophore query using the 2.7 million commercially available screening compounds and the eight features pharmacophore model were used as input. Seven out of eight pharmacophore features had to be addressed by a molecule in order to be selected as a hit. Pharmacophore features F2, F4, F6 and F8 were marked essential. The first conformation of a compound that obeyed all criteria of the pharmacophore model was selected as a hit and the screening automatically continued with the next compound in the database. After clustering and removal of structural analogs, this screen yielded a total of 450 hits.

Next in the *in silico* screening cascade was an energy minimization step. The pharmacophore hit conformation of a compound was energy minimized inside the apo-EphA4 binding site. During this experiment, the protein was kept completely rigid, with the exception of the gatekeeper threonine hydroxyl group. The behavior of each compound was carefully monitored during this energy minimization step. Compounds that retained hydrogen bond interactions with the protein hinge and gatekeeper residue were selected for further investigation. However compounds that lost these key interactions, or were even expelled from the protein binding site, were discarded.

Figure 5.3: *In silico* screening cascade. The numbers of compounds processed in each step are indicated.

The application of the energy minimization step yielded a set of 195 molecules that conformed to the criteria above. The molecules were subjected to a docking experiment using GOLD. The apo-EphA4 crystal structure, in which we introduced one mutation, was used for these docking studies. The sidechain of Lys^{653} was mutated into a methionine sidechain in order to overcome unwanted bias towards ionic interactions with the protonated nitrogen atom of the lysine. Ten poses for each compound were generated. Scoring and ranking of poses in the docking of fragments and drug-like molecules is still very challenging[197] and therefore the poses were visually inspected. Here, the placements of the docked compounds were compared with the energy minimized poses of the same compounds generated in step 3 of the screening cascade.

Compounds that maintained key interactions (i.e., hydrogen-bonding) with the site were selected for further investigation. However, compounds that were docked with binding modes which deviated from the energy minimized poses, but were involved in hydrogen bonding with the hinge and gatekeeper residues, were also selected. This resulted in the identification of 80 compounds which were selected for purchase. Based on the availability, pricing and delivery times of these scaffolds, 27 compounds (including close structural analogs of *in silico* hits that were unavailable) were purchased at different vendors. The inhibitory properties of these compounds on the EphA4 kinase activity were tested in a phosphorylation inhibition assay. This assay resulted in the identification of 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine (**24,** Fig.5.4), a direct analog of *in silico* hit **25** which was unavailable for purchase. Compound **24** exhibited 80% inhibition of the EphA4 kinase activity at 500 μM.

Hit optimization

Results from the pharmacophore screen and docking experiments were used to optimize the affinity for EphA4. The putative binding mode of **24** was derived from the *in silico results* of the structural analog **25** (Fig 5.4A). The proposed binding mode of **1**, after pharmacophore placement, is shown in Figure 5.4B. The 5,6,7,8-tetrahydroisoquinoline aromatic center matches the aromatic feature F4, while the nitrogen atom matches the H-bond acceptor feature F3. The vector projecting from this nitrogen atom intersects feature F2 that was marked essential. Pharmacophore feature F5 is probed by the pyrazolo NH group and the vector projecting from this hydrogen-bond

donor extends towards the essential feature F6. The other pyrazolo nitrogen atom matches the acceptor/donor feature F7 and projects its hydrogen atom towards the essential feature F8. Upon docking of **25** using GOLD, however, a different binding mode for this scaffold was observed (Fig. 5.4C). In this pose, the scaffold is both rotated and flipped compared to the pharmacophore placement. Thus, the 5,6,7,8-tetrahydroisoquinoline nitrogen atom is involved in hydrogen bonding with the threonine hydroxyl group of the gatekeeper. The pyrazolo NH group forms a hydrogen bond with the carbonyl oxygen atom of Glu^{700}, while the other pyrazolo nitrogen atom is engaged in hydrogen bonding with the Met^{702} backbone NH group. The NH2 group of 4 forms a fourth hydrogen bond with the backbone carbonyl oxygen atom of Met^{702}. Superimposing this binding mode with the pharmacophore model reveals that seven out of eight pharmacophore features are hit, including the marked essential features F2, F6 and F8.

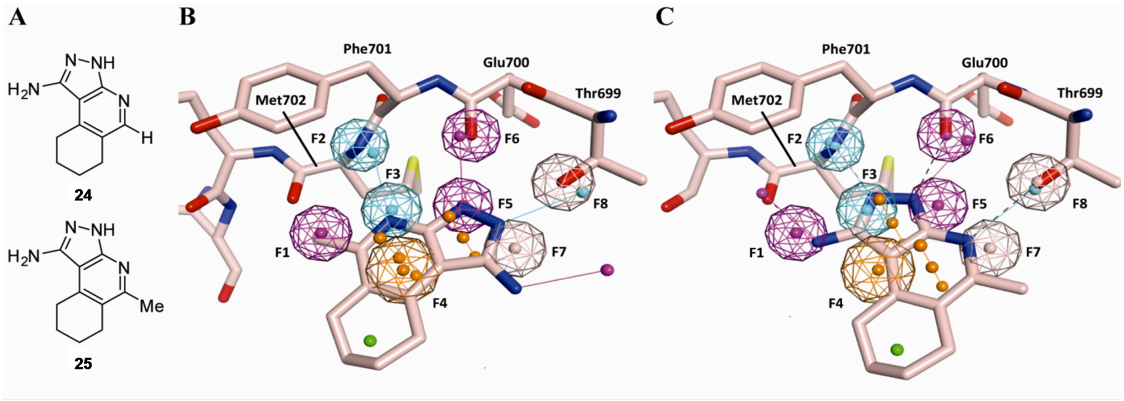


Figure 5.4. Comparison of alternative binding poses of *in silico* hit **25** during pharmacophore placement and after docking. Projecting vectors of hydrogen bond donors (purple), hydrogen bond acceptors (magenta) and aromatic centers (orange) are shown. **A**, structures of **24** and *in silico* hit **25**. **B**, placement of **25** in the pharmacophore screening model. The EphA4 hinge and gatekeeper residues are shown for illustration. **C**, placement of **25** in EphA4 after GOLD docking. Dashed lines indicate hydrogen bonds. The pharmacophore model is added as illustration.

Inspection of Figure 5.4C reveals that the essential aromatic feature F4 is not contacted by **25** in the GOLD docking pose, as such, the scaffold would have been discarded during a pharmacophore screen which required that all eight pharmacophore features had to be addressed. However, a pose similar to the pharmacophore placement of **25** was not observed during the GOLD docking experiment. The putative GOLD-derived binding mode indicates that the methyl group of **25** is pointing towards the hydrophobic back-pocket of EphA4 (Fig. 5.5A). In the Abl1 kinase complex **II**, this hydrophobic back pocket is occupied by the 2-chloro-6-methylphenyl moiety of dasatinib (Fig. 5.5B).

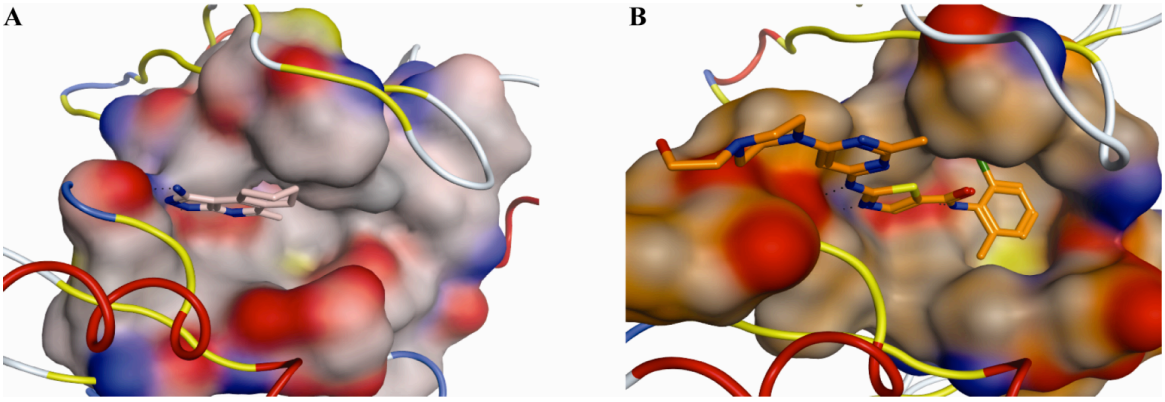


Figure 5.5: **A**, GOLD-derived putative binding mode of **25** in EphA4 (PDB ID: 2HEL). **B**, binding mode of dasatinib in Abl1 (complex **II**, PDB ID: 2GQG).

Upon superimposition of both complexes (image not shown), the methyl group of **25** is located near carbon atom 1 of the 2-chloro-6-methylphenyl moiety of dasatinib. Thus, if the proposed binding mode of **25** is correct, growing into this hydrophobic back-pocket on the methyl side of the fragment should be allowed and result in increased affinity for EphA4. To confirm this hypothesis, hit exploration experiments were performed. Selected compounds were screened in single point measurements at 200 μM

compound concentration for inhibition of EphA4 kinase activity. The inhibition data indicated that introducing a phenyl substituent at position 5 of the 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine scaffold improved affinity (compound **26,** Fig. 5.6). Therefore, synthesis of a series of 5-substituted analogues was performed. Among the compounds tested, a set of three compounds completely inhibited EphA4 activity exhibiting IC_{50} values between 2 and 8 μM. Therefore these compounds were submitted to structure determination via X-ray crystallography. Soaking experiments with compound **27** (IC_{50} 4.5 $\pm$ 0.5μM) yielded a 2.11 Å X-ray complex structure which is described below.

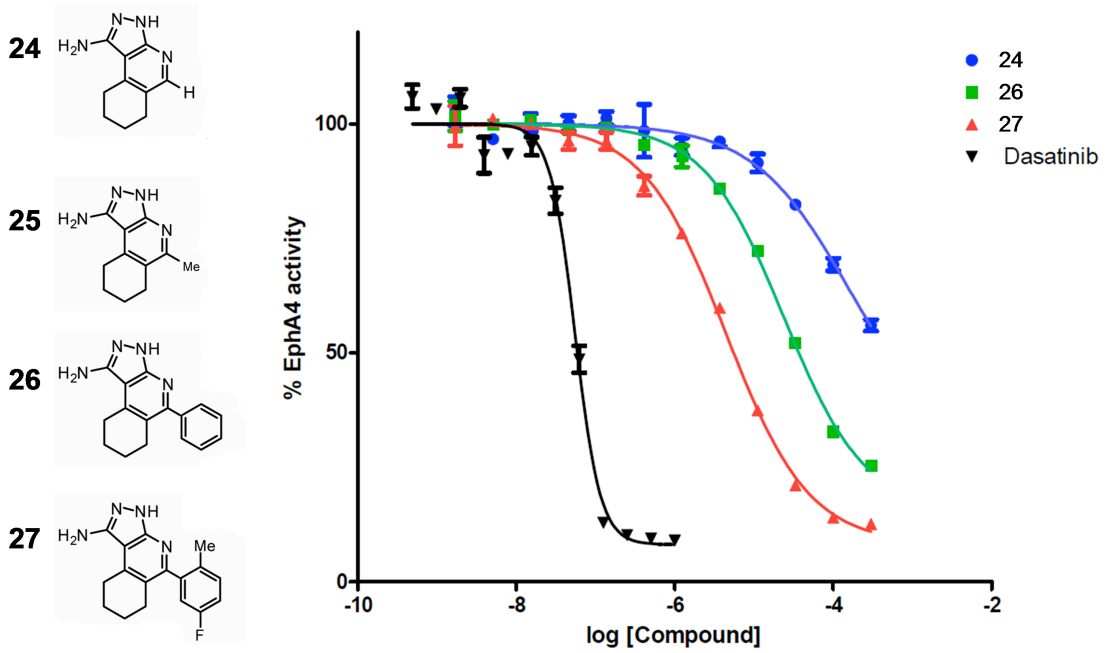


Figure 5.6: Dose dependent inhibition of recombinant EphA4 kinase activity. The structures of compounds **24-27** are presented on the left. On the right, the enzyme inhibition curve of compounds **24**, **26**, **27** and dasatinib are shown. The curves represent the mean ± Standard Error of Mean (S.E.M) of the experiment performed in triplicate.

Crystal structure of the kinase domain of EphA4 in complex with compound 27

To confirm the predicted binding mode of these scaffolds, efforts to soak compound **27** into the available crystal of the EphA4 kinase domain were undertaken. Crystals of apo EphA4 (kinase domain aa 606-846) were grown and the structure was elucidated at 1.7 Å (*c.f.* Chapter 4). Crystals of the apo protein were soaked in 30 mM **27** to obtain the complex, flash frozen and used to collect diffraction data to a resolution of 2.12 Å. The structure of the complex was solved using molecular replacement with the apo protein serving as the search model. The refined structure has been deposited at the protein data bank under the access code: 2XYU.

Alignment of the structures of the apo kinase domain and the complex with **27** indicates that the structure of the kinase is unchanged (r.m.s deviation of 0.179Å over 228 Cα atoms). As predicted, compound **27** makes 4 hydrogen bond contacts within the hinge region of the kinase domain, between the tetrahydroisoquinoline core and Met^{702N}, Met^{702O}, Glu^{700O} and Thr^{699OG1} (Fig. 5.7). The 5-fluoro-2-methyl-phenyl substituent makes extensive van der Waals interactions with Lys^{653}, Glu^{670}, Met^{674}, Ile^{683} and Ile^{697}. The αC helix is in the ATP proximal position allowing the formation of the conserved Lys^{653}-Glu^{670} salt bridge while the activation loop is in the "DFG-in" conformation, both of which are indicative of a kinase in the active state.[167, 198]

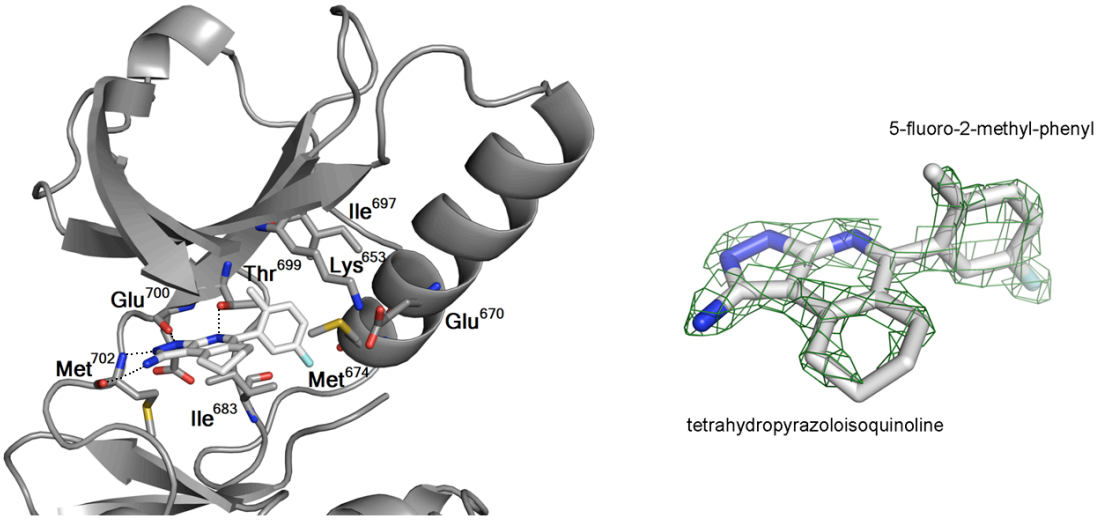


Figure 5.7: Structure of the complex of EphA4 with **27**. The protein backbone is shown as a ribbon diagram. The compound and key residues of the protein with which it interacts are shown in a bond representation. The simulated annealing omit-map Fo-Fc electron density map contoured at 2.0 σ of **27** is also presented in mesh format. The residues involved in hydrogen bonds are shown in bold.

Selectivity profiling

Compound **27** was screened against a panel of 124 protein kinases at 10 μM compound concentration. It fully inhibits the all Eph kinases that were assayed, including EphA4 (Fig. 5.8). However, there is only significant inhibitionof a few other protein kinases (including BTK, Lck, and YES1).[199] Therefore, **27** seems to be a reasonably selective Eph kinase inhibitor, supporting the use of this scaffold in targeted EphA4 inhibition.

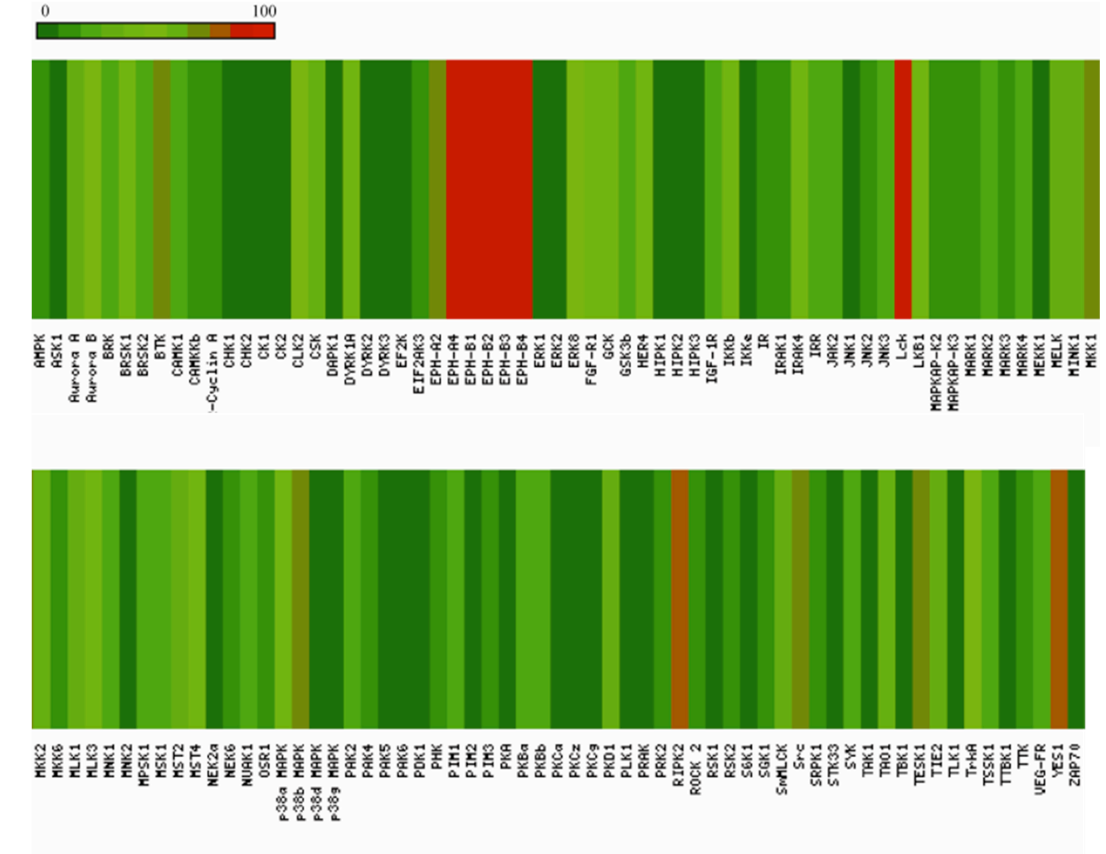


Figure 5.8: Heat map of % inhibition of **73** against 124 protein kinases.

Discussion

In this work we describe the fragment-based discovery of a novel EphA4 hinge binding compound. Using the structural information derived from crystal structures of two ligands bound to related kinases, a mixed pharmacophore model for the binding site of EphA4 was constructed. *In silico* screening efforts, using a virtual library containing commercially available compounds, led to the identification of a 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine fragment i.e. compound **24**. Subsequent optimization of this scaffold, by growing into the kinase hydrophobic back-pocket, resulted in the identification of an EphA4 inhibitor with 4.5 μM (IC_{50}) activity. The binding mode of **27** was elucidated *via* X-ray crystallography and confirmed the predicted binding mode of this scaffold (PBD ID: 2XYU). Selectivity profiling against 124 protein kinases[199] revealed that this compound seems to be a moderately selective kinase inhibitor.

The potential applicability of this scaffold was recently confirmed with the disclosure of a patent and publication by the Merck group.[200, 201] The patent describes the [3,4-]isoquinolin-l-amine scaffold as an inhibitor of EphA4 (Fig. 5.9) for the treatment of diseases regulated by the EphA4 RTK signaling, such as neurological and neurodegenerative disorders and cancer. Meanwhile, the publication presents an extensive study, both in vitro and in vivo, of several compounds including 2 closely related to compound **27** (Fig. 5.9). The reported IC_{50} values of compounds **28** and **29** were respectively of 1.5 and 2.1 μM which are consistent with the data presented in this work. Altogether, this data underscores the potential of our scaffold as a possible new drug for the treatment of cancer and neuronal injuries by inhibition of the EphA4 receptor tyrosine kinase.

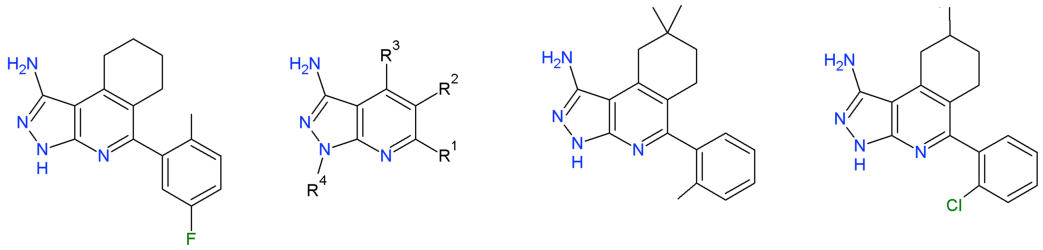


Figure 5.9: Structures of compound **27**, the scaffold and compounds **28** and **29** from the patent and publication by researchers from Merck.

Chapter 6

Biophysical characterization of fragments discovered with TINS

Major drug discovery efforts for novel targeted therapies of various human malignancies focus on the human kinome. Indeed, modulation of phosphorylation signaling pathways by small molecules has demonstrated considerable clinical efficacy in the treatment of devastating disorders such as cancer. Here we present the biophysical characterization of fragments interacting with the EphA4 receptor. Fragments discovered previously with the Target Immobilized NMR Screening approach, were characterized using Surface Plasmon Resonance (SPR) and [^{15}N,^{1}H]-HSQC experiments. We successfully developed a functional SPR assay for EphA4, binding affinities could be determined for several fragments and competition experiments with ATP were undertaken. Our results establish the utility of biophysical methods in the characterization of fragment screening hits.

Introduction

From the approval of the first kinase inhibitor, imatinib[45], for the treatment of chronic myelogenous leukemia (CML) in 2001 to the approval of vandetanib[46] in early 2011, protein kinases have developed into one of the most important target classes in oncology drug discovery. However, studies revealed that most approved kinase inhibitors have limited selectivity and target multiple kinases.[47] EphA4, a receptor tyrosine kinase normally only expressed in the developing brain, was recently found to be highly overexpressed in tumor cells of cutaneous T-cell lymphomas (CTCL),[180] prostate cancer and pancreatic cancer [58, 181, 182] suggesting that it might be a promising target for pharmacologic intervention. Furthermore, it has been suggested that EphA4 inhibition might be beneficial for the treatment of spinal cord injuries.[60, 202]

Over the last decade, fragment based drug Discovery (FBDD) has been recognized as a powerful strategy to develop novel, small molecule drugs.[105, 203] FBDD screens small libraries (1,000 - 20,000 compounds) of so-called drug "fragments" that are typically described by a "rule of threes"[28] (Ro3, Mr < 300 Da, cLogP <3, H-bond donors < 3, H-bond acceptors < 3, number of rotatable bonds < 3 and TPSA (total polar surface area) < 60 $Å^2$) for binding to the target. The crucial step in this drug discovery process is the reliable identification of the initial fragments that typically interact only weakly with the receptor due to their small size. Several biophysical methods such as NMR,[126,27] protein crystallography[204] and surface plasmon resonance (SPR)[205] have been employed to identify such weak protein ligand interaction.

We have developed an NMR ligand screening approach, called Target Immobilized NMR screening (TINS) and have used the method to screen a fragment library for binding to the kinase domain of the EphA4 receptor (Chapter 3). We wished to compare the results of TINS and biochemical methods to those from alternative biophysical techniques such as Surface Plasmon Resonance (SPR) and [^{15}N,^{1}H]-HSQC (Heteronuclear Single Quantum Correlation) experiments. Thus the binding of all fragments discovered in Chapter 3 was studied by SPR. In addition, protein observed NMR was used to study the binding of fragments to the EphA4 KD under conditions similar to those used for crystal soaking (Chapter 3 and 5).

Materials and method

Protein purification for Surface Plasmon Resonance studies

The enzymatically biotinylated AVI-KD of EphA4 was purified as described in Chapter 3, while the 6His fusion KD of EphA4 was purified as described in Chapter 4.

Immobilization for Surface Plasmon Resonance studies

All SPR studies have been performed using a Biacore T-200 instrument (GE Healthcare). Sensor Chip CM5 and NTA and the amine coupling kit were purchased from GE healthcare. All buffers were freshly prepared and filtered through 22 μm membranes.

Immobilization of the 6His-Kinase Domain of EphA4:

Capture was performed at 15 °C by injecting the His kinase fusion protein over a Ni-NTA chip previously charged with 0.5 mM $NiCl_2$ at 10 μl/min for 5 min. The chip was regenerated with 1 M imidazole, followed by 350 mM EDTA and 2 M NaCl at 20 μl/min. Injections were repeated until saturation or until the desired level was reached. The running buffer used in all subsequent binding and inhibition experiments was composed of 20 mM Tris Buffer, 200 mM NaCl, 1mM $MgCl_2$, 0.05% Tween 20, pH 7.5 with 5% of DMSO. On average, 3000 response units (RUs) of 6His-EphA4 were immobilized.

Immobilization of the biotinylated Kinase Domain of EphA4:

Coupling of NeutrAvidin (PierceNet Thermo Scientific) to the CM5 surface was performed using the amine coupling kit from GE healthcare according to the instructions of the manufacturer. In brief, NeutrAvidin was first dissolved (10 mM HEPES pH 7.5, 50 mM NaCl and 0.01% Tween 20) at a concentration of 50 μM and then diluted to 5 μM in 10 mM Na-acetate buffer at pH 4.8. The functional groups of the CM5 chip surface were first activated by injecting a mixture of carbodiimide (0.5 M) and N-hydroxysuccinimide (0.5 M), immediately followed by injecting the NeutrAvidin at a flow rate of 5 μl/min. The remaining activated carboxyl groups on the surface were blocked with 50 μM sodium hydroxide. Approximately 7500 response units (RUs) of NeutrAvidin were immobilized. The running buffer used in all subsequent binding and inhibition experiments was composed of 20 mM Tris Buffer, 150 mM NaCl, 2 mM $MgCl_2$, 0.05% Tween 20, pH 7.5 with 5% of DMSO. The biotinylated protein was immobilized at a concentration of 5 μM for 5 min at 10 μl/min resulting in approximately 13000 RU of protein immobilized. As a reference, biotin was immobilized in the same conditions, reaching an immobilization level of approximately 80 RU.

Analysis of fragment-EphA4 KD interactions by SPR biosensor

Equilibrium binding experiments:

A 100 mM stock solution of each compound in 100% DMSO was diluted into the running buffer to obtain an aqueous fragment solution at 750 μM concentration

containing 5% DMSO. For compound **27** from chapter 5, the stock solution of 500 μM in 100% DMSO was diluted to 25 μM. For titration experiments, samples were injected in a series of 6 concentrations with a maximum of 750 μM or 1500 μM with the exception of compound **27** for which titrations were run with a maximum of 50 μM. These solutions were injected over the protein sensor surface for 25s at a flow rate of 30 μl/min and after every 25 injections the reference compound **27** was injected as a control at a single concentration of 25 μM. In addition, titrations of compound **27** were performed at the beginning and at the end of each plate, in order to assess the stability of the surface. These control measurements were performed to check both the stability and ligand-binding activity of the immobilized proteins during a screen. SPR experiments using the 6His fusion KD were performed at 15°C and those using the biotinylated KD were performed at 20 °C. For the 6His fusion KD screen, regeneration of the surface between subsequent binding experiments was achieved by washing the surface extensively (30 s) with the running buffer plus 10% DMSO. Sensor responses were taken during each injection after a contact interval of 30 s. Report points at 5 seconds before the end of the injection were used for analysis.

Data analysis:

All sensorgrams were processed using a double-referencing method.[206] First, the binding response from the reference surface was subtracted from the binding response from the surface containing EphA4 to correct for bulk refractive index changes. Second, the response from an average of the blank injections was subtracted to remove any systematic artifacts observed between the reaction and reference flow cells.

Assuming linearity between molecular weight and refractive index, the percentage of active protein immobilized could be estimated from the saturation response of the reference compound (**27**) according to equation 1.

$$R_{\max} = \frac{MW_{comp}}{MW_{prot}} \times R_{prot} \qquad (1)$$

R_{max} is the saturation signal determined for the reference, R_{prot} is the signal measured for the amount of immobilized protein [RU], and MW_{comp} and MW_{prot} are the respective molecular weights of reference compound and protein.

Data was fitted using GraphPad Prism version 5.00 for Windows, GraphPad Software, La Jolla California USA. The K_D value could be estimated using equation (2) derived from the Langmuir adsorption isotherm.

$$K_D = \frac{R_{\max} \times C}{R} - C \qquad (2)$$

R_{max}, R, and C correspond to the normalized saturation response of the reference compound, the normalized response of the test compound, and the concentration of the test solution, respectively.

Competition experiments:

The competition experiments were performed by sequentially injecting the solution of the test compound (fragment), the reference compound at saturation concentration i.e. ATP (500 μM), and a mixture of fragment and reference compound each for 30 seconds. The test compounds were compound **27** (2 μM), compound **44** and **7**

(750 μM). A qualitative comparison of the signals observed for test compound, reference compound and mixture was then performed. For competition, experiments the running buffer was 20 mM Tris Buffer, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween 20, pH 7.5 with 5% of DMSO. In addition, a titration of ATP was performed. Samples were injected in a series of concentrations with a maximum of 1000 μM.

The evaluation of competitiveness in this experiment required specific data analysis. Fragments displaying high affinity relative to the competitor and binding to the same binding pocket as ATP, showed a signal for the mixture (ATP + fragment) that was identical to the signal observed for the fragment alone. In the case of low affinity fragments, where saturation was difficult to reach, the response that would be observed if the compound is competitive with ATP for binding to EphA4 (i.e. binds to the ATP binding pocket) is given by the sum of the fractional occupancies (FO) of ATP and the fragment. The fractional occupancies of the mixture of the 2 compounds can be calculated using equations 4 and 5.[207]

$$FO(A) = \frac{1}{1 + \frac{K_D(A)}{C_A}\left(1 + \frac{C_B}{K_D(B)}\right)} \qquad FO(B) = \frac{1}{1 + \frac{K_D(B)}{C_B}\left(1 + \frac{C_A}{K_D(A)}\right)} \qquad (4)$$

K_D(A) and K_D(B) are the equilibrium dissociation constants of the competing compounds, and C_A and C_B are the respective concentrations. In the case of competition between compounds of similar affinity and equal concentration, the occupancy of the binding sites by one of the components is heavily influenced by the presence of the other. From the estimation of the occupancies, the expected signals of the 2 compounds could be estimated using equation 5.

$$R_{observed} = FO(A) \times R_{\max}(A) + FO(B) \times R_{\max}(B) \qquad (5)$$

in which R_{max}(A) and R_{max}(B) are the saturation responses of the 2 compounds and FO(A) and FO(B) are the respective fractional occupancies.

Protein production for NMR studies

The plasmid pET28b containing the kinase domain of EphA4 (amino acids 606-846) was transformed in *E. coli* BL21 RP competent cells. The cells were incubated overnight at 37ºC on a LB/Kan/Cam (0.05 g/l Kanamycin, 0.034 g/l Chloramphenicol) plate. A single colony was inoculated in 50 ml M9 minimal medium supplemented with 0.3 g/l $^{15}NH_4Cl$ and 0.05 g/l Kanamycin, 0.034 g/l Chloramphenicol in a 250 ml flask. Five mL of the pre-culture was inoculated in 0.5 L M9 minimal medium (0.3 g/l, $^{15}NH_4Cl$, 0.05 g/l Kanamycin, 0.034 g/l Chloramphenicol). The cultures were grown at 37ºC, shaking at 250 rpm to an OD_{600} of 0.6 before induction with 0.1 mM IPTG. At the same time the temperature was lowered to 18ºC and growth was continued overnight before harvesting. A cell pellet corresponding to 1 liter of culture was resuspended in 30 ml of lysis buffer (50 mM Sodium phosphate pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% Tween 20, 10% glycerol and 0.01% beta-mercapto-ethanol) supplemented with a protease inhibitor cocktail (complete EDTA-free) from Roche, lysed using a French press and centrifuged at 37,000 x g for 45 min at 4°C. The supernatant was applied to a 5 ml HisTrap HP column (GE Healthcare) with a peristaltic pump at 4°C for 1 hour (Binding buffer: 20 mM Tris-HCl pH 7.6, 500 mM NaCl and 5 mM imidazole). After extensive washing, the recombinant protein was eluted with binding buffer plus 500 mM imidazole.

The 6His fusion EphA4 was incubated with 20 mM ATP and 20 mM $MgCl_2$ in order to fully activate it. After an hour at 4°C, the protein was diluted 10 fold in 10 mM HEPES pH 7.5, 10 mM NaCl and applied to a 1 ml HiTrap Q column (GE healthcare) that had been pre-equilibrated with the same buffer. The fractions containing the EphA4 kinase domain, which did not bind to the column, were then concentrated to 5 ml and applied to a Sephacryl S-200 HR column (HiPrep 16/60; GE Healthcare) equilibrated in 10 mM HEPES, 500 mM KCl and 1 mM DTT.

NMR sample preparation

The 6-His fusion KD was concentrated using a Centricon-10 (Amicon). All protein samples contained identical buffer condition 20 mM HEPES, 150 mM NaCl, 2 mM DTT, pH 7.2, in 3% D_2O. The sample for recording the apo protein spectra contained 80 μM ^{15}N-6His-KD. The sample for pH titration contained 100 μM of protein and the pH, ranging from 5.8 to 6.2, was adjusted with 0.1 M HCl or 0.5M NaOH. The samples for compound titration contained 100 μM of protein in buffer and 100 mM stock solutions of compounds **27**, **3** and **5** were prepared in d_6-DMSO.

NMR Spectroscopy

All NMR spectra were recorded at 298 K on a Brucker DMX600 spectrometer. [^{15}N,^{1}H]-HSQC were obtained with spectral widths of 28 ppm (^{15}N) and 16.0 ppm (^{1}H). Data was processed with Topspin 2.1 and analyzed with Sparky.[208]

Results

Development of SPR biosensor assay for EphA4

Initially, the same immobilization strategy as the one used for TINS studies (*c.f.* Chapter 3), *i.e.* immobilization *via* the SNAP affinity tag was used for SPR biosensors studies. As a reference, the SNAP protein was immobilized in a separate channel. To establish the validity of the SPR assay, a well-characterized ligand was first titrated. As compound **27** was biochemically validated and the 3D structure was available (*c.f.* Chapter 5), this compound was chosen as a reference However, the titration of compound 27 did not exhibit a concentration dependent response (Data not shown) suggesting that the SNAP mediated immobilization strategy was not suitable for SPR studies.

As an alternative,, we attempted to immobilize the KD of EphA4 via the oligo histidine tag on an Ni-NTA sensor chip. Purified 6His fusion KD was captured on the sensor chip and a titration of compound **27** was performed (Fig. 6.1). As seen in Fig. 6.1A, binding of **27** to the metal affinity immobilized KD was dose responsive and saturated at an appropriate level. The binding affinity of compound **27** was determined using equation (2) derived from the Langmuir adsorption isotherm (Fig. 6.1A). Analysis of the titration data indicates that compound **27** binds with an affinity of 1 μM which is comparable with the affinity measured by biochemical assay (IC_{50}= 4.5 μM, *c.f.* Chapter 5). The IC_{50} of compound **27** is slightly higher than the K_D, this was expected given that the inhibitor competes with ATP for the same binding site in the activity assay.

The kinetic association (k_a) and dissociation (k_d) constants for binding of **27** to immobilized EphA4 could be determined by fitting the data to a 1:1 binding model. The fitting yielding values of k_a =2 x 10^5 $M^{-1}s^{-1}$ and k_d=0.5 s^{-1} (Fig. 6.1B). The resulting kinetically determined dissociation constant $K_{D,}$ of 2 μM matches the equilibrium affinity determined from the Langmuir isotherm.

The binding curve in Figure 6.1 indicates that the saturation response for compound **27** is approximately 18 RU. Using the molecular weight of the immobilized protein and reference compound **27** (37,282 Da and 296 Da respectively) and the amount of immobilized protein (R_{prot}), the theoretical saturation response for 100% active protein is estimated to be 29 RU according to equation (1). Comparing the theoretical and experimental saturation responses, the percentage of immobilized protein that was active for ligand binding was estimated to be about 63%. Other known ligands such as ATP and ADP were assayed, but no binding was observed and therefore compound **27** was used as a reference for monitoring the stability of EphA4 during the SPR assay.

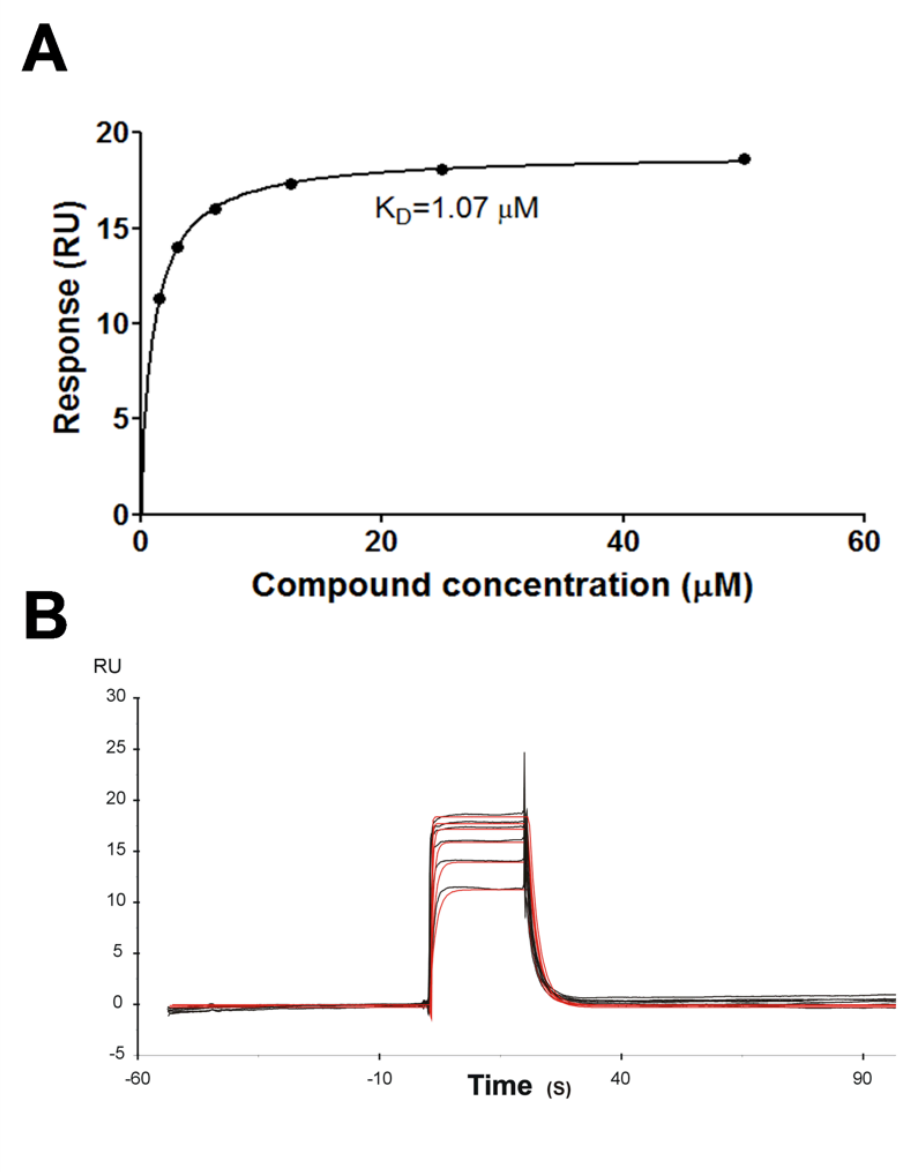


Figure 6.1: SPR characterization of compound **27** binding to EphA4 6His-KD. A) Representative saturation curve of compound **27** binding. The response signal is plotted against the concentration of compound **27**. B) Sensorgrams measured at the following concentrations: 1.56, 3.125, 6.25, 12.5, 25 and 50 μM. The sensorgrams were processed using the double referencing method. The red lines overlaid on the experimental black curves result from the fit of the experimental data with the Biacore T200 evaluation software (version 1.0) using the mathematical function of a kinetic 1:1 binding model.

Biophysical characterization by SPR of hits from SNAP fusion KD TINS screen

Initially, compounds from the TINS screen of the SNAP fusion KD were selected and biochemically characterized (*c.f.* Chapter 3). We wished to compare the biochemical results to the SPR results. The SPR experiment was carried out on the 6His KD of EphA4 immobilized on an NTA chip and 134 compounds were assayed at a single concentration of 750 μM. In order to compare the results from the SPR assay to the biochemical assay and the ligand screening results from TINS, compounds were categorized in groups for more clarity (Table 6.1). Compounds considered as hits in the TINS screen (for more details refer to Chapter 3) and demonstrating biological activity were classified in group I. Group II contained hits from the TINS experiment but that did not show biological activity. Fragments that exhibited inhibition in the enzymatic assay but were not considered as hits in TINS screen belonged to group III. Finally, negative controls i.e. compounds that did not show either binding in TINS experiment nor biological activity, were classified in group IV.

Table 6.1: Comparison of hits from TINS, biochemical assay and SPR. In the TINS screen of SNAP KD fusion, compounds were selected as binders when the T/R was lower than 0.4 (Chapter 3, fig. 3.5).For the enzyme inhibition assay, compounds were considered as inhibitors when the PIN was above 10 % (Chapter 3, table 3.1). Regarding the SPR experiments, fragments were identified as hits when the signal exceeded 3 times the SD and lower than the maximal theoretical response (see below).

	Group I	Group II	Group III	Group IV
	+ TINS +E.I	+ TINS -E.I	-TINS +E.I	- TINS -E.I
SPR positives	6	63	1	3
SPR negative	6	48	1	6
Total of compounds	12	111	2	9

E.I.: Enzyme inhibition assay

Fragments were identified as hits in the SPR assay when the signal exceeded 3 times the noise and the response was lower than the maximum theoretical response at saturation (i.e. binding was 1:1). Ill-behaved compounds were eliminated if they showed evidence of super stoichiometric binding, precipitation and/or a sensorgram that indicates non-equilibrium behavior.[209] Using these criteria, 73 out 134 compounds were considered as hits in this SPR assay.

Half of the compounds from group I and a majority of compounds from group II were categorized as binders in the SPR assay suggesting a correlation between the SPR experiment and the TINS assay. As the detection window was quite small (11 RU in the present experiment), some compounds may not have been positive in the SPR assay due

to insufficient sensitivity. Moreover, the extremely high sensitivity of TINS (up to ~20 mM K_D)[149] allows identification of many compounds that may not be detected in the SPR assay which has an upper limit of 2-5 mM under ideal conditions.[210]

Of the compounds from group II that were hits in TINS and biochemically negative, 63 were confirmed as hits in the SPR experiment suggesting that either these compounds bind too weakly to inhibit or that they bind at a non-biologically productive site. Out of the two compounds from group III, one fragment that was negative in TINS but positive in the bioassay was detected as positive in the SPR experiment. This compound was deemed to be a false negative for TINS. Interestingly, the other compound which was negative in both TINS and SPR, showed activity in the biochemical assay, possibly suggesting that the compound was a false positive in the biochemical assay. The 3 compounds in group IV categorized as positive in the SPR experiments are probably false positives.

In order to gather more biophysical information, the 73 well behaved compounds from the single concentration SPR assay were subsequently titrated. Compounds exhibiting low response at 750 μM (i.e. less than 0.5 x the R_{max} observed for compound **27**) were titrated using a concentration range from 46 μM to 1500 μM. Meanwhile, compounds exhibiting a higher response in the single concentration assay were titrated using a concentration range from 23 to 750 μM. Out of these 73 compounds, 12 presented a clear concentration-dependent response. The binding affinity of these fragments was estimated by SPR to range from 0.75 to 3 mM (Appendix C). Out of these 12 fragments, 3 were from group I and 9 from group II. Compounds from group III and IV were titrated but did not show any concentration dependent response.

For compounds **2**, **3** and **8** from group I *i.e.* positive in the TINS screen and biochemical assay, the binding affinities were estimated to be 800±400, 900±400, 1100±700 μM (Fig. 6.2). The measured responses were analyzed using a non-linear method derived from the Langmuir model and points where the binding was less than 5% of the maximum response are considered to be highly inaccurate and were ignored.[211] The large uncertainty derives from the fact that binding could not be saturated.

It is known that the NTA surface has the inherent problem of leaching of the immobilized protein which leads to a "drooping" baseline.[212] Consequently, it was difficult to determine with precision the binding affinities for many of the compounds analyzed. Despite a highly variable level of immobilization (between 1500 and 3000 RU, also frequently encountered with this immobilization strategy) and the leaching off the surface, useful data was still achieved and a rough estimate of the binding affinity could be determined for a few compounds. Nonetheless, a more stable system with a larger response window would clearly be advantageous.

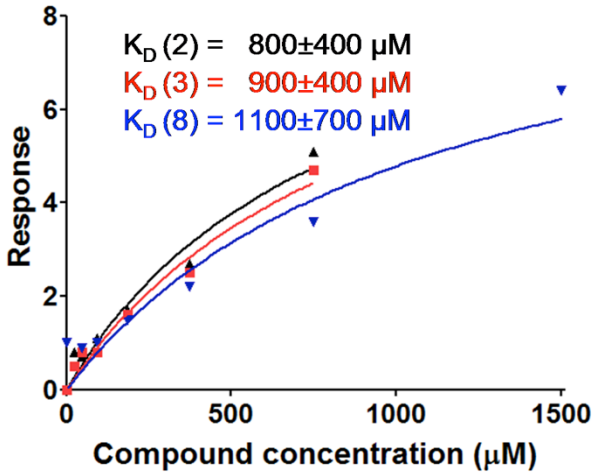


Figure 6.2: SPR analysis of fragment binding to EphA4 KD immobilized via the 6His tag. Representative saturation curves of compounds **2**, **3** and **8** from Group I compounds. The response signal is plotted against the concentration of compound.

Biophysical characterization by SPR of hits from the Biotinylated KD TINS screen

To further characterize the hits selected during the TINS ligand screen of the biotinylated KD, binding to EphA4 was characterized by SPR using the same immobilization approach as used in TINS. The kinase domain of EphA4 was enzymatically biotinylated at the AVI tag and captured on a biosensor chip *via* immobilized NeutrAvidin. To validate the assay, the binding of compound **27** was monitored and results were compared to those from the immobilization using the 6His tag (Data not shown). The K_D of **27** for biotin immobilized EphA4 KD was estimated to be ~ 1.9 μM and the percentage of active protein was approximately 95%. This binding affinity for compound **27** is similar to the one obtained with the 6His KD immobilization on the NTA chip (1.9 μM for the present immobilization compared to 1-2 μM for the immobilization on NTA chip) while the percentage of active protein is significantly higher. Further comparison of the two immobilization methods is presented below.

A total of 62 compounds was selected as hits in the TINS experiment using biotinylated EphA4 (c.f. Chapter 3) and assayed for binding at a concentration of 750 μM *via* SPR technology. Of these 62 compounds, 51 were positive for binding in the SPR experiment, while 11 did not show binding, giving an 82% validation of the selected TINS hits. The well-behaved compounds (as previously defined) amongst the 51 binders were further analyzed to determine their equilibrium binding affinity by titration experiments. Reproducible, concentration dependent curves were obtained for 8 compounds. Examples of the experimental binding curves are shown in Figure 6.4 for compounds **7** and **44** and in Appendix D for the remaining compounds. Binding

affinities of these 8 fragments were estimated to range from 0.4 to 2.6 mM. The binding affinity for compounds **7** and **44** was estimated to be 650±65 μM and 460±60 μM respectively.

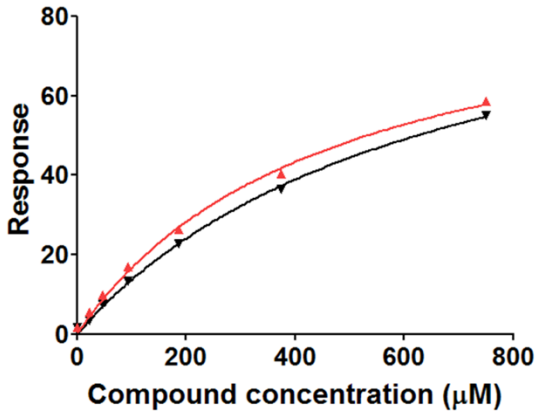


Figure 6.3 SPR analysis of fragments binding to EphA4 KD immobilized *via* NeutrAvidin. Saturation binding curves of compounds **7** and **44**. The response signal is plotted against the concentration of each compound. Compounds 7 and **44** are represented in black and red respectively.

Comparison of the two SPR immobilization approaches

We wished to compare the relevance of the two SPR approaches: immobilization of 6His fusion KD on the NTA chip and ligand capture of biotinylated EphA4. A total of 13 compounds were tested in both SPR assays, 8 were positive in both assays and 3 were negative in both assays. Meanwhile 2 compounds were detected as binding in the SPR assay on the biotinylated KD but not in the 6His SPR assay. One possible explanation resides in the window for ligand detection. In the immobilization mediated with NTA

chip, the level of immobilization was significantly lower than the one in the ligand capture of biotinylated KD (3,000 RU vs. 13,000 RU). This resulted in a smaller window for ligand detection: ~11 RU for 6His KD immobilization vs. ~100 RU for biotinylated KD immobilization. It is likely that the SPR signal for these 2 compounds was close to the noise in the 6His KD immobilization, thus they appeared as false negatives. In addition, compound **32** was titrated using the 2 different immobilization procedures and the binding affinities obtained were very similar. (K_D ~900±400 μM in the SPR screen of the 6His fusion KD and K_D~900±600 μM in the SPR screen of biotinylated KD).

Competition experiments

Competition binding experiments allow one to differentiate artifactual ligand binding from real binding and can provide insight into the binding site on the target. Given the inability to obtain crystal structures of screening hits bound to EphA4, the information from competition binding experiments could be used to enhance the reliability of computational docking experiments.

Among the compounds titrated using SPR technology, compounds **7** and **44** presented the highest affinity relative to the other fragments. Therefore, these two compounds were assayed for competitive binding with the natural ligand ATP using SPR. Since immobilization via enzymatic biotinylation proved superior, this approach was used to conduct the competition experiments. It was necessary to first demonstrate well-behaved binding of ATP to the immobilized KD. Since kinases typically bind ATP as a complex with Mg^{2+}, $MgCl_2$ was included in the running buffer and binding of ATP was

now clearly observed. This result suggests that the reason for the lack of ATP binding to the SNAP-KD may have been the absence of Mg^{2+}. ATP was subsequently titrated in order to determine the binding affinity (Fig. 6.4). Binding of ATP to immobilized EphA4 KD was characterized by k_{on} and k_{off} rates that were faster than the maximal detectable rates of the Biacore instrument, as has been previously observed for other kinases[213] (Fig. 6.5A). Hence, the K_D value was calculated from the equilibrium response of the ATP injections at various concentrations. The fit of the Langmuir isotherm for ATP (Fig. 6.4B) gave a K_D value of 270 ±15 μM.

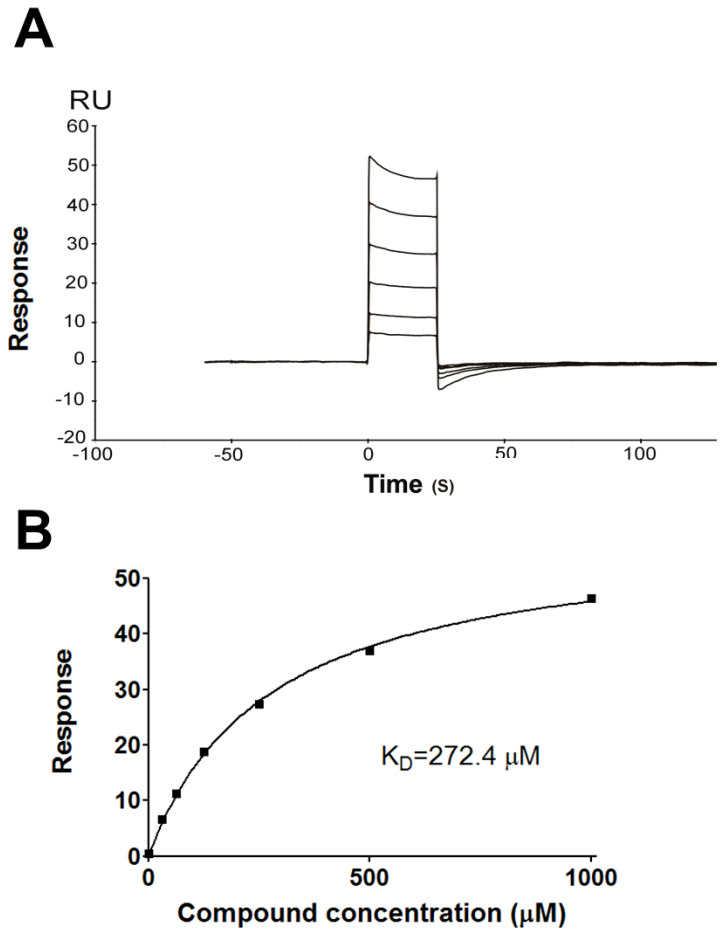


Figure 6.4: A) Sensorgram of ATP binding to immobilized EphA4 KD measured at the following concentrations: 35.25, 62.5, 125, 250, 500 and 1000 μM. B) Representative saturation curve of ATP binding.

Two compounds that simultaneously bind to the KD would be expected to generate an additive binding response in SPR.[214] In contrast, a compound that binds to the ATP binding pocket would compete with ATP and therefore give a non-additive response in SPR. Based on the crystal structure of compound **27** in complex with EphA4, which showed the compound bound in the ATP binding pocket, this particular compound was chosen to validate the SPR competition binding assay (Fig. 6.5). Compound **27** was tested in the presence and absence of ATP at a saturating concentration. The SPR response observed in the presence and absence of ATP is the same for **27**. This result is what is expected when a 10 fold more potent ligand competes with a weaker ligand, i.e. complete displacement of the weaker ligand.

Figure 6.5 summarizes the results of the competition experiments that were performed with fragments **7** and **44** versus ATP. In these cases, we observe that the signal of the mixture (black bars) is different from the individual signals of compounds **7** and **44** (white bars) or ATP (dark gray bars). It has been shown that in the case of competition between compounds of similar affinity, the occupancy of the binding site lies between the individual occupancies and the sum of the 2 individual occupancies.[207] The expected signal of the two compounds could be estimated using equations 4 and 5 in the material and method section.

The response for the mixture (ATP + fragment) was calculated in the case of fragment competing with ATP for the same binding site (R_{calc} competition, light gray bars). To be consistent, the response in case of noncompetition (i.e., in the case of independent binding sites for each of the compounds) was also calculated (R_{calc}

noncompetitive, striped bars). When there is no competition, the expected signal is the sum of the two individual responses observed for the fragments.

For compounds 7 and 44, the experimental responses for the mixtures ATP + fragment are closer to R_{calc} competition than to R_{calc} non competition (Fig. 6.5) Since the experimental binding response is closer to the expected result for competitive binding within experimental error, the data is strongly in favor of competition and suggest that the binding site of compounds **7** and **44** overlaps with ATP. These results imply that compounds **7** and **44** both bind at the ATP binding site.

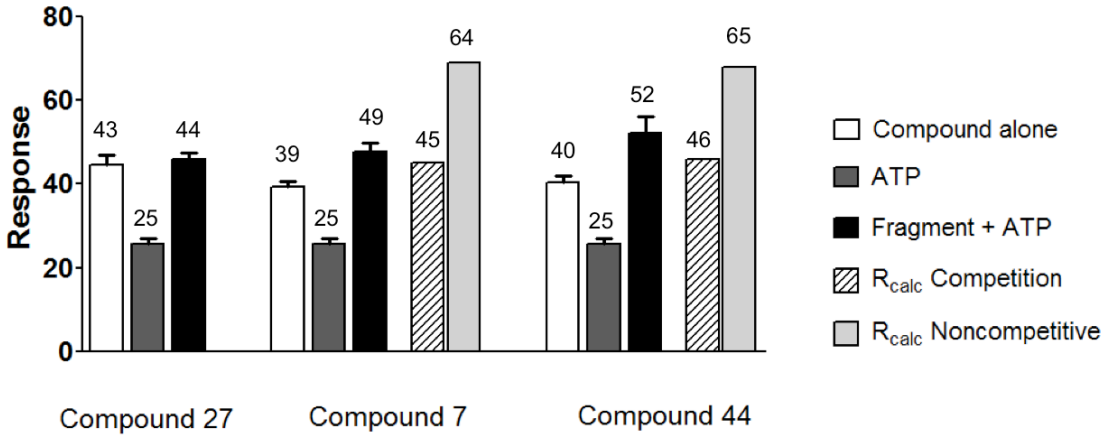


Figure 6.5: SPR analysis of competition binding experiments. The measured response for each compound is indicated by the white bars: compound **27** (2 μM) compounds **7** and **44** (750 μM), the response for ATP (500 μM) by dark gray bars and the response for the mixture (ATP (500 μM) + compound (2 μM for compound **27** or 750 μM for compounds **7** and **44**) is shown by black bars. The calculated signal for the mixture (ATP+**7**, ATP+**44**) expected for fractional occupancy of the same binding pocket using the measured affinities and known concentrations, is shown by the striped bars. The calculated signal for the mixture (ATP+**7**, ATP+**44**) expected for different binding pockets (i.e. non-competitive binding) is shown in light gray bars. The mean value is indicated above its respective bar.

NMR characterization of apo-EphA4

Specific binding of a small molecule to a protein will result in concentration dependent, observable changes of the NMR resonances of the protein.[27] The most convenient manner to observe these spectral changes is to monitor the amide chemical shifts of a ^{15}N labeled protein. Chemical shifts of both ^{1}H and ^{15}N amide atoms are recorded in the [^{15}N,^{1}H]-HSQC (Heteronuclear Single Quantum Correlation) spectrum. In addition, when the sequential assignment of a protein is available, analysis of chemical shifts perturbation data affords rapid access to low resolution structural data to characterize the ligand binding site.[27, 215] [^{15}N,^{1}H]-HSQC NMR titration experiments were used to characterize binding of fragments to EphA4 at a pH equivalent to the one in crystallographic soaking experiments. While the resonance assignment of the kinase domain of EphA4 is not available, spectra of the kinase domain of EphB2 are of high quality, resulting in a complete backbone resonance assignment for this protein.[133] The [^{15}N,^{1}H]-HSQC spectrum of uniformly ^{15}N labeled EphA4 protein in the apo-form is presented in Fig. 6.7. The spectrum displayed good resonance dispersion, however the spectrum was sufficiently different from that of EphB2 that the assignments could not be readily transferred. Even in the absence of assignments, chemical shift perturbation studies can be used to characterize binding of a fragment to a protein.

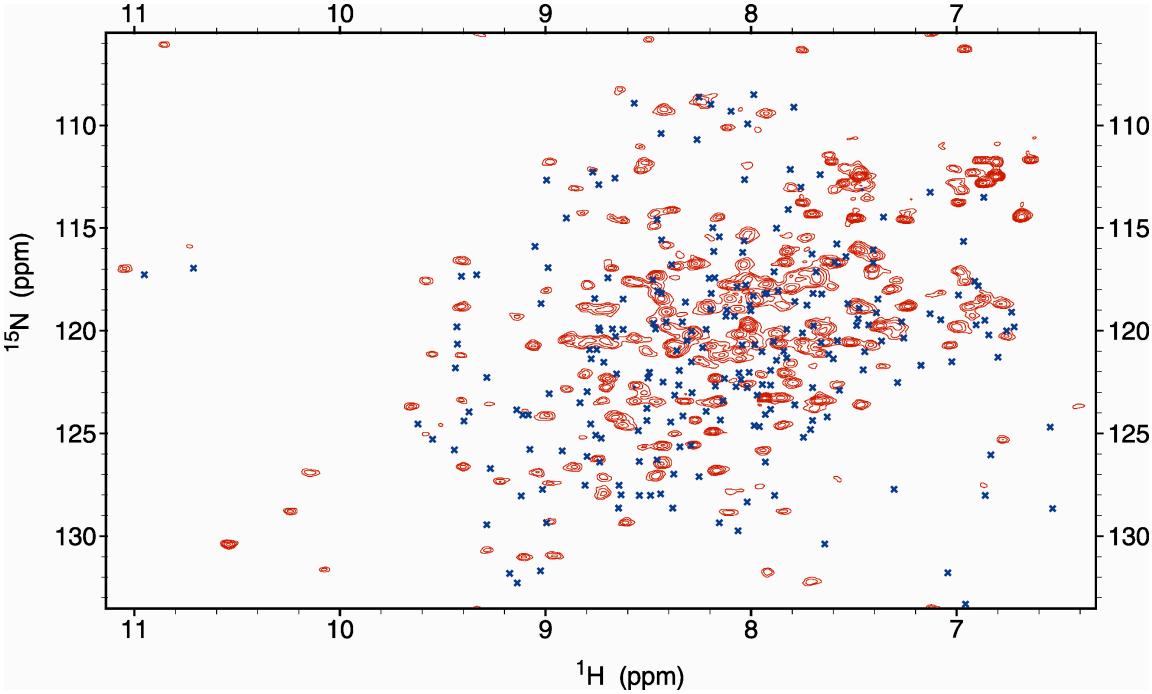


Figure 6.7: [^{15}N,^{1}H]-HSQC spectrum of 80 µM ^{15}N EphA4-KD. The resonance frequencies of the [^{15}N,^{1}H]-HSQC spectrum of EphB2-KD are represented by blue crosses. No clear correlation between the spectra is observed

pH titration of ^{15}N EphA4 KD

Since it is well known that even small changes in pH can cause changes in the [^{15}N,^{1}H]-HSQC spectrum, it was important to first assess the sensitivity of EphA4 to pH. To do so the pH was titrated in small increments while monitoring the protein by [^{15}N,^{1}H]-HSQC NMR. Spectra were acquired at pH 5.8, 6.0, and 6.2 in order to investigate the possible effect of small changes to the pH used for crystal soaking (pH 6) potentially caused by titration of ligands to high concentration. The superimposition of the [^{15}N,^{1}H]-HSQC spectra at various pH values is presented in Appendix E, while the region showing peaks used for analysis of ligand binding is presented in Figure 6.8. There were no significant chemical shift differences observed in this pH range.

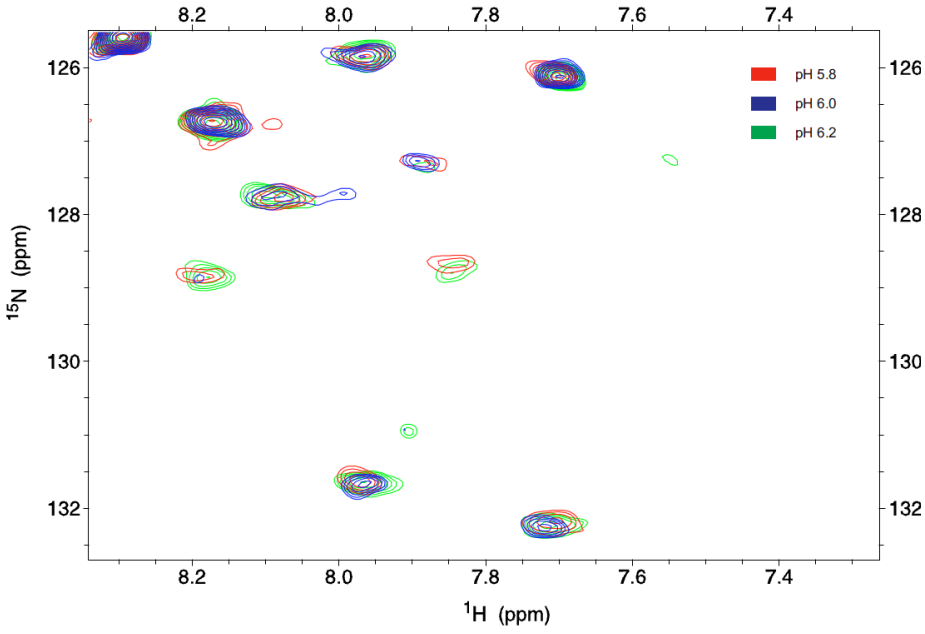


Figure 6.8: pH titration of ^{15}N EphA4 KD. A region of the overlaid [^{15}N,^{1}H]-HSQC spectra of 100 µM EphA4 KD is recorded at pH 5.8, 6.0 and 6.2. This region shows peaks used for analysis of ligand binding.

Titration of compound 27

Initially the well characterized ligand **27** was used to assess the effect of compound binding on the [^{15}N,^{1}H]-HSQC spectrum of EphA4. Addition of compound **27** to ^{15}N EphA4 KD resulted in significant changes in the [^{15}N,^{1}H]-HSQC spectrum. The resonances exhibiting the largest changes upon compound **27** binding are shown in Figure 6.9 while the full spectra are presented in Appendix F. We observe that upon addition of compound **27**, peaks 1-3 exhibit significant changes in resonance position that are primarily indicative of slow exchange (that is $k_{off} << \Delta(\delta bound-\delta free)$). On the other hand, peak 4 showed both a gradual decrease in peak intensity and concentration dependent chemical shift change, suggesting intermediate exchange (that is $k_{off} \approx \Delta(\delta bound-\delta free)$). As the resonance assignment of EphA4 is not available, these 4 peaks, which report on ligands binding at the ATP site, are simply referred to by number. Since the binding of **27** to EphA4 could readily be detected, it was deemed worthwhile to investigate the binding of fragment screening hits using this approach.

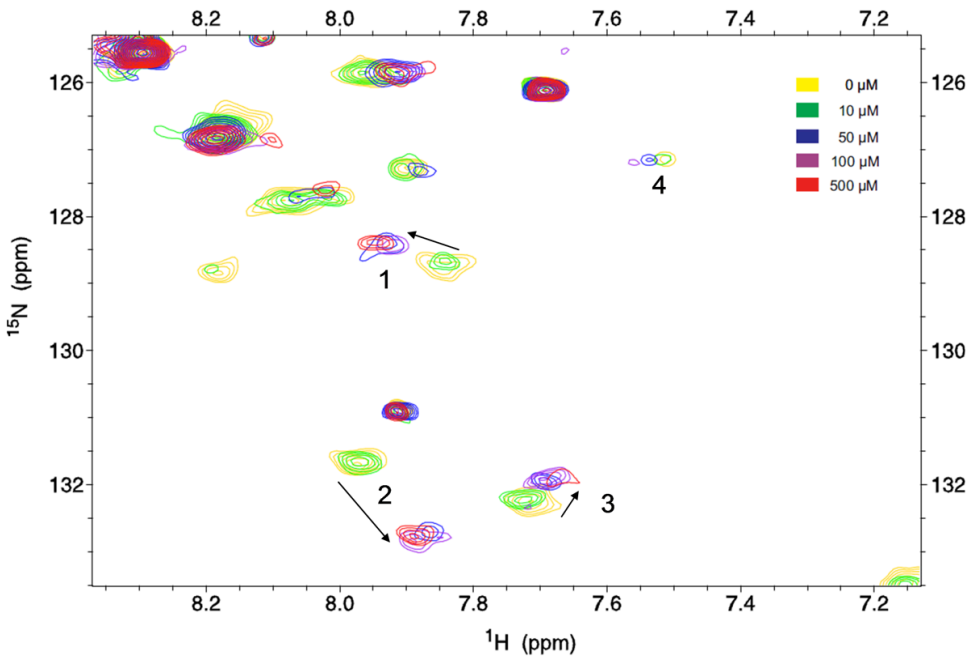


Figure 6.9: NMR analysis of compound **27** binding to ^{15}N EphA4 KD. A region of the overlaid [^{15}N,^{1}H]-HSQC spectra of 100 µM of ^{15}N EphA4-KD with increasing concentration of compound **27** is shown. Peaks whose resonance position changes with increasing concentration of compound **27** are indicated with arrowheads and numbers.

Fragment binding to ^{15}N EphA4 KD

Perturbations in the spectrum of EphA4 KD were monitored upon addition of different fragments. For clarity, we focused on a small region of the spectrum containing peaks reporting on binding to the ATP pocket as determined from the titration of **27** (Fig. 6.10). We selected compounds **3** and **5** from the TINS screen of the SNAP fusion KD since they were also active in the enzyme inhibition assay yet negative for crystallography. Figure 6.11, shows that upon addition of compound **3**, peaks 2 and 3 are shifted, meanwhile in presence of compound **5** peaks 1 and 3 are shifted. The same peaks are shifted in the presence of compound **27** (Fig 6.11B) and these peaks are not related to

a change in pH (Fig. 6.11C). It is likely that compounds **3** and **5** bind in a similar site as **27** *i.e*. in the ATP binding site.

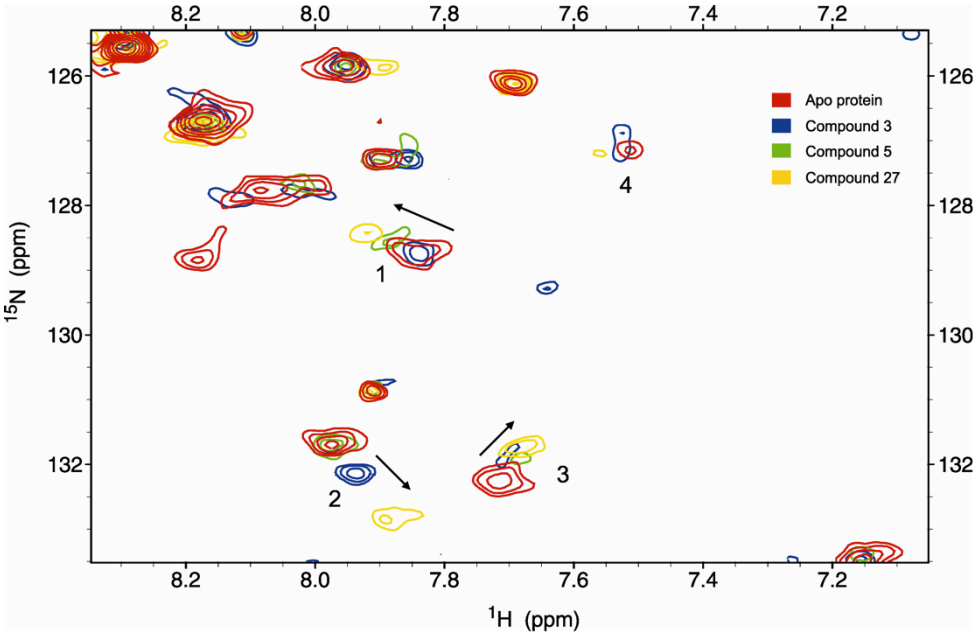


Figure 6.10: NMR analysis of ligand binding to EphA4-KD. A region of the overlaid [^{15}N,^{1}H]-HSQC spectra of 100 µM apo ^{15}N EphA4-KD and in the presence of the indicated compounds is presented. Compound **3** (blue)and compound **5** (green) were added to 2 mM and compound **27** (yellow) was added to 100 µM. Peaks whose resonance position changes with increasing ligand concentration are indicated with arrowheads and numbers.

Discussion

The initial plan of this collaborative research project was to develop potent, specific inhibitors of EphA4 kinase activity using both computational and experimental approaches. As we have seen, the computational approach has led to a reasonably potent and specific compound (**27**) for which a crystal structure has been elucidated bound to EphA4. Surprisingly, the experimental approach of using a biophysical assay for ligand screening and confirming with a biochemical assay has failed to yield a single compound for which a crystal structure could be determined. Moreover, several fragment screening hits were selected for elaboration studies but these did not lead to compounds with similar potency as **27**. There are a number of possible explanations for these observations. One is that the form of the protein that was screened using TINS was not biologically relevant. A second is that the biochemical assay was not functioning as a proper filter to differentiate real from artifactual binding. The experiments described in this chapter aimed to determine whether either of these explanations were correct.

Surface Plasmon Resonance (SPR) was employed to provide orthogonal biophysical characterization of the compounds binding to EphA4. To obtain high quality SPR data, the target protein has to be immobilized in a way that maintains its activity. The SPR characterization was conducted with 2 different recombinant proteins (6His fusion KD and biotinylated AVI fusion KD) using 2 different mechanisms of protein immobilization (via Ni-NTA surface and NeutrAvidin immobilized on CM5 surface). As compound **27** was biochemically validated and the binding mode was confirmed by crystallography (*c.f.* Chapter 5), this fragment was chosen as a reference compound for

all SPR studies. The kinetics for compound **27** binding to EphA4-KD immobilized via both methods were fit to a 1:1 binding model yielding kinetic constants k_a =2 x 10^5 $M^{-1}s^{-1}$ and k_d=0.5 s^{-1}, and a K_D ~ 1.5 ±0.5 μM. This binding affinity is comparable with the affinity measured by biochemical assay (IC_{50}= 4.5 μM, *c.f.* Chapter 5). The data suggested that both immobilization protocols resulted in an SPR assay capable of characterizing small molecules binding to immobilized EphA4 KD. However, immobilization of biotinylated KD presented several advantages when compared to immobilization of the 6His fusion EphA4 KD: (i) the level of immobilization is significantly higher yielding a greater window to study compound binding, (ii) 95% of the calculated binding capacity is retained and (iii) immobilization *via* the biotin NeutrAvidin interaction is more stable.

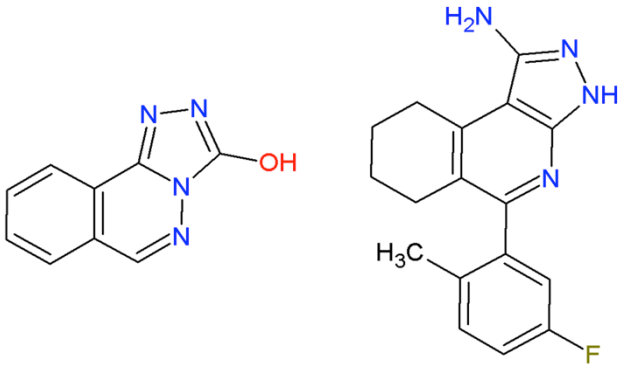
Compounds selected as hits from the TINS screen were characterized by SPR using both Ni-NTA and biotinylated KD mediated immobilization. Overall, there was a good correlation between the results from TINS screen and the SPR characterization. Approximately 55% of the hits from the TINS screen on the SNAP-KD protein were determined as binders in the SPR assay, while 82% of TINS hit from the screen of the biotinylated KD were validated. Several compounds that were positive for binding in TINS may have been negative in SPR experiments due to the limited sensitivity of the method. It has been shown that dynamic range limitations of SPR can lead to false negatives.[100] In SPR experiments, the observed response is directly proportional to the ratio $[L]/K_D$. The lower limit of $[L]/K_D$ leading to a detectable response in is in the order of 0.2.[216] Since concentrations higher then 0.75-1 mM of ligand tend to lead to poor behavior in the Biacore instrument, the practical cutoff for ligand affinity is ~4 mM.

Compounds with weaker affinity than this will appear as false negatives in the SPR experiment.

Subsequently well behaved fragment with reproducible curves that showed clear dose–response and the same qualitative sensorgram shape for each concentration were selected for titration. The binding affinities of several fragments could be estimated and ranged from 0.4 to 3 mM, approximately. However, the binding affinity could not be obtained for a majority of the compounds. There are several possible explanations for this observation. One is that the fragments were too weak and thus saturation could not be reached. In some case compounds may have exhibited promiscuous binding *via* a concentration-dependent aggregation mechanism.[209] In addition, it may also have been possible that the stringent curation of the dose–response curves may have resulted in some false negatives.[210]

The fragments with highest affinity and best SPR behavior were assessed for competition binding using ATP. The two compounds, **7** and **44,** were originally selected from the two different TINS screen of EphA4 KD. Compounds **7** and **44** had reasonable ligand efficiencies of 0.3 and 0.25 (ΔG/# of heavy atoms) respectively and were likely to bind in the ATP binding site. In principle these compounds could form starting points for hit to lead studies to generate novel EphA4 inhibitors. Interestingly, compound **7** has a very similar structure to compound **27** (Figure 6.11) and the affinity is similar to precursors of 27 that lack the attached derivatized benzene ring. The X-ray crystallographic structure of **27** showed that it binds the hinge region of the ATP binding site. Based on the similarity to compound **27**, it seems likely that compound **7** exhibits a

closely related binding mode to EphA4 KD, that is, it forms hydrogen bonds to the backbone atoms of the hinge residues.



7 **27**

Figure 6.11: Structures of compound **7** (left) and compounds **27** (right).

In order to investigate whether the low yield of fragment structures obtained with X-ray crystallography was related to the difference in pH between screening and biochemical characterization on the one hand and crystallography on the other, protein observed NMR was employed. [^{15}N,^{1}H]-HSQC spectra of EphA4-KD were acquired in the presence of different fragments at the same pH as for crystallography *i.e* pH~6.0. The experiments conducted on fragments **3** and **5**, indicate that the change in pH does not preclude compound binding for these 2 fragments. These compounds were submitted to X-ray crystallography but exhibited weak diffraction i.e. higher than 7 Å.

The lack of crystal structures of compounds from the TINS screen, in addition to the odd behavior of the compounds in the biochemical assay, lead us to question the relevance of the TINS screen. In this chapter, SPR and [^{15}N,^{1}H]-HSQC NMR were used to confirm interaction of the compounds with EphA4 KD. Moreover, compounds **7** and

44, each originating from a different TINS screen, were shown to bind to the ATP binding site. The similarity of the structure of compound **7** and **27** suggests that the same compounds found computationally have also been found experimentally, a reassuring and exciting result.

Chapter 7

General conclusions and Perspectives

Due to their involvement in a large number of pathologies, protein kinases present an important pharmacological interest. Among this family, EphA4 has been recognized to be differentially expressed in various human tumors. The aim of the research project that forms the subject this thesis was to develop potent, specific inhibitors of EphA4 kinase activity using both computational and experimental approaches.

Target Immobilized NMR screen applied to the EphA4 kinase domain

As described in **Chapter 2**, nuclear magnetic resonance spectroscopy (NMR) is a powerful tool that can provide important information at each and every step of drug development. From ligand screening to hit validation, NMR is commonly used for characterizing the structure and molecular dynamics of target or ligand molecules. During structure-based lead optimization, NMR gives insight into the structural and dynamic properties of the target-ligand complex. NMR screening methods range from the detection of signal from protein observed to ligand-observed methods. Recently, we developed a screening method called target immobilized NMR screening (TINS), in which a target and a reference protein are immobilized on solid media and binding of fragments to the immobilized proteins is monitored by NMR. Among the advantages of this method is the small amount of protein required as TINS uses a single sample of the target to screen an entire fragment library.

In **Chapter 3**, a fragment library was initially screened against the Kinase Domain (KD) of EphA4 using the TINS technology. In order to insure the physiological relevance of ligands discovered using TINS, it was important to retain full enzymatic activity of

EphA4 upon immobilization. Therefore a number of immobilizations chemistries were investigated and we determined that oriented immobilization mediated by an affinity tag was required in order to obtain a functional immobilization of EphA4 KD. Subsequently, TINS was applied to the SNAP fusion KD of EphA4 using two different T_2 relaxation periods of 80 ms and 2 ms. These screens resulted in a 6.8% and 3.9 % hit rate respectively, with a final set of 180 fragments which were selected for further characterization using a kinase inhibition assay.

Subsequent SPR studies described in **Chapter 6** yielded an alternative means of immobilization via interaction between enzymatically biotinylatedEphA4 and streptavidin covalently bound to a bead. Since the biotin mediated immobilization yielded more biochemically active enzyme, the same technique was employed to perform an additional TINS screen. The results from this second screen were compared to those from the SNAP-KD screen. Despite the fact that TINS screens typically exhibit good reproducibility, little correlation was observed between the two screens and the TINS profiles were different. At this point it was not clear where the variability came from. Moreover, the protein was successfully crystallized using different protein batches and yielded several structures as described in **Chapters 4 and 5** and below. Thus variation in the quality of the protein preparation was not a likely source of the unexpected results. As a functional SPR assay was developed, the technique was employed to investigate the source of this variability.

In **Chapter 6,** two compounds (compounds **7** and **44)**, each originating from a different TINS screen, were characterized with SPR and shown to bind to the ATP binding site. The structure of compound **7** was found to be similar to compound **27** for

which the crystal structure was determined bound to EphA4 KD **(Chapter 5)**. Based on the structural similarity of compounds **7** and **27**, it seems that the same chemical scaffold has been found both computationally and experimentally, suggesting that both approaches can be used to generate EphA4 KD inhibitors.

Crystallography

Many strategies are available for obtaining crystals of protein-ligand complexes. The most resource-effective method of obtaining the structure of a protein–ligand complex is by soaking the ligand of interest into apo protein crystals. Therefore, the first step was to obtain diffracting crystals of apo-EphA4. **(Chapter 4)** Crystallographic experiments were undertaken with the support of the Nederlands Kanker Instituut (NKI-AVL) in Amsterdam and the structure of the native EphA4 kinase domain was elucidated (PDB ID: 2Y6M). In addition, the crystal structure of EphA4 in complex with dasatinib was solved (PDB ID: 2Y6O) and revealed a binding mode closely related to the one exhibited by Src family members, c-Src and Lyn. Analysis of the two structures revealed a hydrophobic back-pocket in the ATP-binding site of EphA4 which was unknown before. The access to this hydrophobic pocket is governed by the gatekeeper residue.[172] It is likely that targeting this hydrophobic pocket during drug design could lead to inhibitor specificity.

Discovery of a series of potent and specific inhibitor

In parallel, a computational approach toward EphA4 was performed by the Ph.D. candidate Oscar van Linden in the group of Dr. de Esch in Amsterdam (**Chapter 5**). Using the structural information derived from crystal structures of two ligands bound to related kinases, a mixed pharmacophore model for the binding site of EphA4 was constructed. An *in silico* screening procedure led to the identification of a 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine fragment. Optimization of this scaffold by growing into the kinase hydrophobic back-pocket discovered in the apo structure of EphA4, resulted in the identification of compound **27.** Soaking experiments were undertaken to obtain structural information and yielded a 2.11Å X-ray structure of the EphA4 – inhibitor complex. The binding mode observed in the crystal structure was in concordance with the binding mode of the scaffold as proposed by the initial *in silico* work (PBD ID: 2XYU). This compound was further biophysically characterized using Surface Plasmon Resonance (SPR) technology in **Chapter 6** and the binding affinity was estimated to be 2 μM. The kinetics for compound **27** binding to immobilized EphA4-KD were fitted to a 1:1 binding model yielding kinetic constants k_a =2.155 x 10^5 $M^{-1}s^{-1}$ and k_d=0.497 s^{-1}, and a K_D ~ 1-2 μM. This binding affinity is comparable with the affinity measured by biochemical assay (IC_{50}= 4.5 μM, **Chapter 5**).

Development of a functional SPR assay for characterization of small molecules binding to EphA4 KD

SPR was employed to provide biophysical characterization of the compounds binding to EphA4. The SPR characterization was conducted with 2 different recombinant proteins (6His fusion KD and biotinylated KD) using 2 different mechanisms of protein immobilization (via Ni-NTA surface and NeutrAvidin immobilized on a CM5 surface). Binding of compound **27** was characterized by SPR using the two different immobilization procedures and was found to be optimal with the biotinylated protein. Immobilization of biotinylated KD had several advantages when compared to immobilization of the 6His fusion EphA4 KD: (i) the level of immobilization is significantly higher yielding a greater window to study compound binding, (ii) 95% of the calculated binding capacity is retained and (iii) immobilization via the biotin NeutrAvidin interaction is more stable.

Biochemical and biophysical characterization of the TINS fragments hits

The hits obtained with TINS were biochemically characterized using an enzyme inhibition assay. Surprisingly, most of the compounds did not exhibit any inhibition of EphA4 activity while only 15% of the fragments exhibited biological activity. Among the fragments that were characterized as inhibitory, most presented peculiar dose response curves with a steep Hill slope and the curves appeared to level at less than 100% inhibition. One possible explanation for the biochemical data was either aggregation or

precipitation of the compounds. However, these explanations were deemed unlikely as the library used for the screen was carefully designed and tested for solubility. However, a moderate number of compounds were found that exhibited some inhibition and presented an interesting chemical scaffold were therefore selected for crystallographic structure determination.

As the soaking experiments successfully yielded a structure of compound **27** in complex with EphA4, compounds originating from TINS were submitted to the same procedure. However, none of the fragments yielded a structure. Most of the crystals that had been soaked with a fragment did not diffract to high resolution and when they did, the experimental electron density map that could be obtained revealed no density for the ligand. As there is a pH difference between the crystallographic conditions and the biochemical assay, this could negatively influence binding of the fragments. Thus protein observed NMR was employed in **Chapter 6** to investigate whether the pH difference could be the cause for the lack of crystal structures. [^{15}N,^{1}H]-HSQC experiments on EphA4-KD were conducted and indicated that the change in pH does not preclude compound binding for the 2 fragments assessed. To gain further insight into the possible sources of the difficulties to obtain protein-ligand structures, SPR characterization of the fragments was performed.

Compounds selected as hits from the TINS screen were characterized by SPR using both Ni-NTA and biotinylated KD mediated immobilization. For simply detecting binding, a good correlation was observed between the results from the TINS screen and the SPR analysis: 55% of the hits from the TINS screen of the SNAP-KD protein and 82% of TINS hits from the screen of the biotinylated KD were validated. Moreover,

compounds with weaker affinity than 4 mM could have been characterized as false negatives in SPR due the dynamic range limitations of SPR **(Chapter 6).** The binding affinities of several fragments could be estimated and ranged from 0.4 to 3 mM. However, for a majority of the compounds the binding affinity could not be obtained. Many factors likely contributed to the inability to determine the affinity including: the fragments may bind too weakly and thus saturation could not be reached, compounds may have exhibited promiscuous binding via a concentration-dependent aggregation mechanism and it may also have been possible that the stringent curation of the dose–response curves may have resulted in some false negatives.

SPR competition experiments using ATP were performed on two compounds presenting the highest affinity. The two compounds, **7** and **44,** were originally selected from the two different TINS screen of EphA4 KD and were shown to bind to the ATP binding site. Compounds **7** and **44** had reasonable ligand efficiencies of 0.3 and 0.25 (ΔG/# of heavy atoms) respectively. These compounds constitute starting points for the generation of more potent EphA4 inhibitors.

Perspectives

Generally, fragments are identified using a biophysical screening method, supported by a biochemical assay to validate their biological relevance. Using only affinity to guide drug discovery has been successful in numerous HTS campaigns in which potent hits are uncovered.[17] However, in our case, most of the fragments did not display biological activity in the biochemical assay even if their binding was demonstrated with both TINS and SPR approaches. It is likely that the biochemical assay used in this project was not sensitive enough for weak inhibitors. As a biological assay is am important tool in a drug discovery project, it is important to carefully develop and select a relevant methodology.

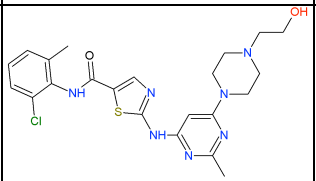
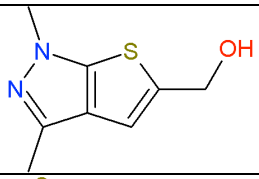
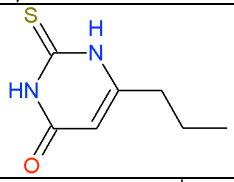
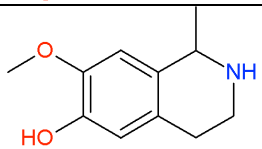
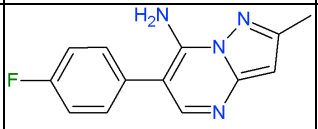
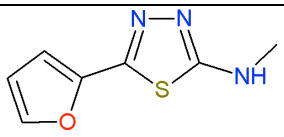
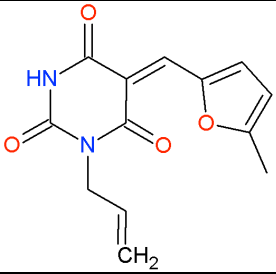
Once hits are identified, their binding needs to be confirmed and the three dimensional structure of the protein-fragment complex should be determined, ideally by X-ray crystallography. Among the approaches available to obtain protein-ligand structures, soaking fragments into existing crystals was chosen in this study. The soaking approach yielded a protein ligand structure for compound **27** but was unsuccessful with the compounds originating from the TINS screen. It is possible that the protein conformation was not optimal for ligand binding of some ligands, or the fragments were not sufficiently ordered to yield observable electron density. As most of the crystals did not diffract to high resolution, it would be interesting to investigate post-crystallization treatments for improving diffraction quality of protein crystals.[217] Furthermore, another approach for obtaining structures of ligand-target complexes, cocrystallization, has been used in drug discovery programs with success.[207] However, cocrystallization is far more

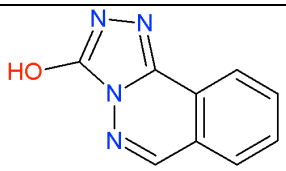
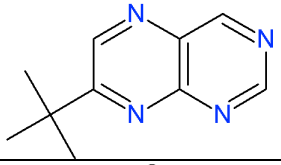
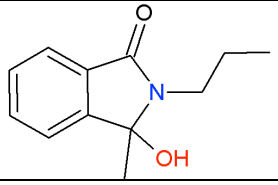
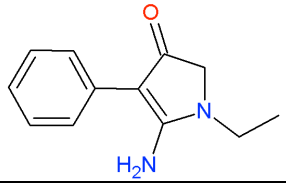
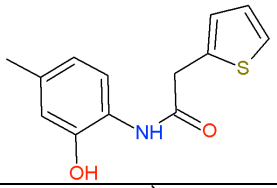
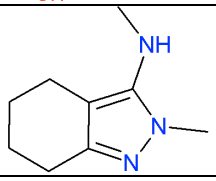
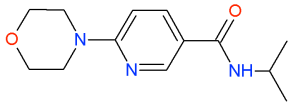
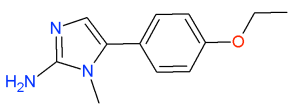
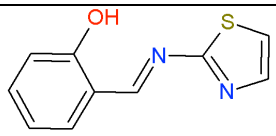
resource intensive and therefore not realistic for the large numbers of fragments assayed in this project. However, in the few cases where the diffraction quality of the fragment soaked crystals was low, it would be interesting to perform co-crystallizations experiments on EphA4 KD with confirmed hits from TINS.

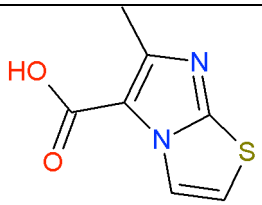
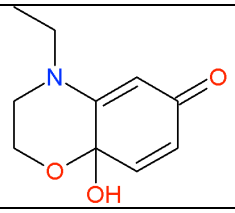
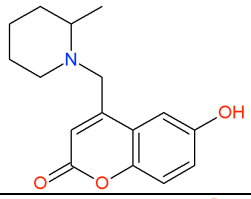
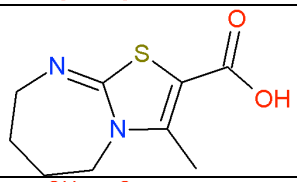
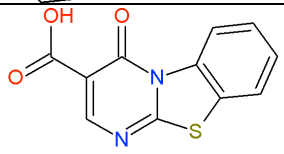
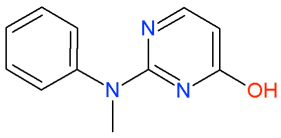
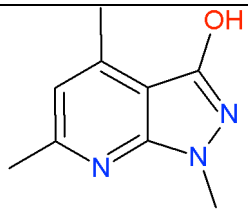
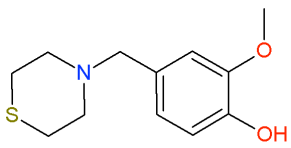
In this study, the *in silico* approach led to the identification of compound **27**, a 6,7,8,9-tetrahydro-3*H*-pyrazolo[3,4-*c*]isoquinolin-1-amine fragment. This compound was demonstrated to be reasonably potent and the binding mode was elucidated via crystallography. Selectivity profiling against 124 protein kinases[199] revealed that this compound appeared to be a reasonably selective kinase inhibitor. Moreover, this scaffold was recently disclosed in a patent and publication by the pharmaceutical company Merck [200, 201] thus confirming the potential of this scaffold for generating EphA4 inhibitors. It is interesting to note that a small academic collaboration discovered the same chemical scaffold as a large pharmaceutical company with, one assumes, considerably less resources. Considering the selectivity of this compound, even though moderate, and the ligand efficiency of 0.35, it represents an interesting starting point to design specific inhibitors. Further improvements of this scaffold could lead to a new drug for the treatment of cancer and neuronal injuries by inhibition of the EphA4 receptor tyrosine kinase.

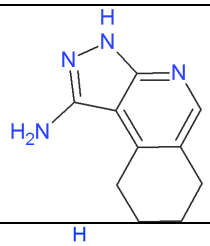
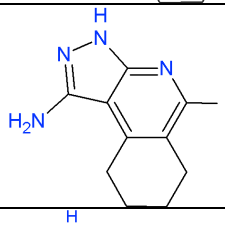
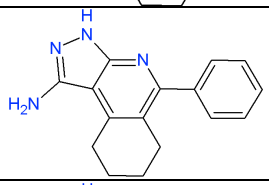
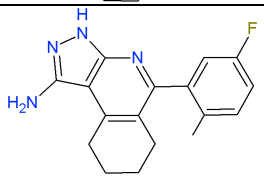
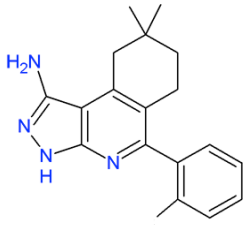
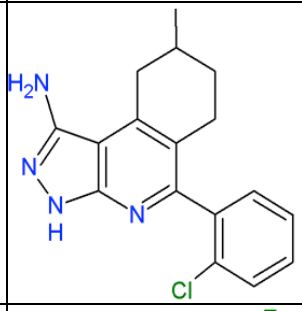
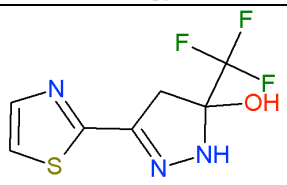
Appendix A: Summary of the compounds studied in this thesis

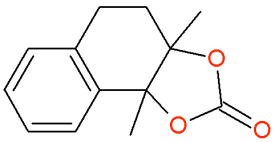
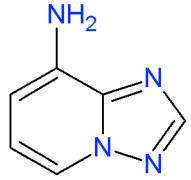
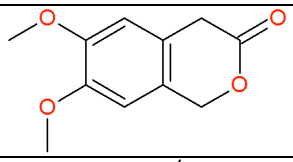
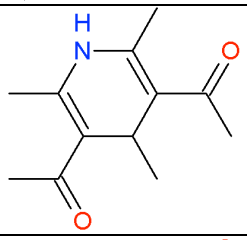
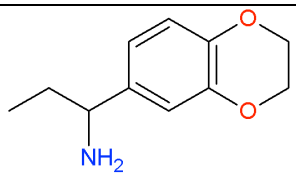
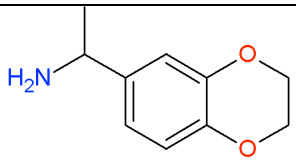
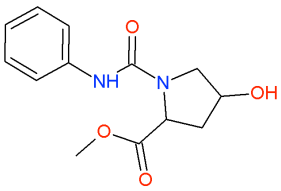
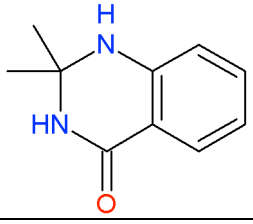
The compounds studied in this thesis are all presented in the table below with the corresponding structure, molecular weight, origin, and when available biochemical data (percentage of inhibition or PIN, IC_{50}), and the K_D obtained with SPR experiments.

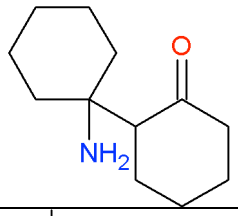
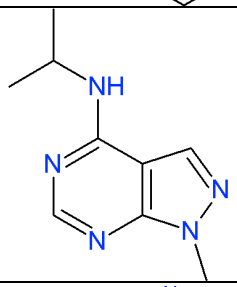
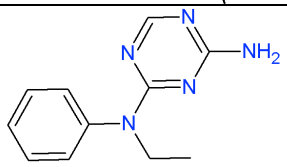
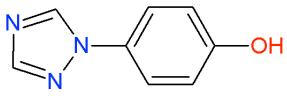
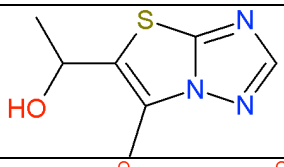
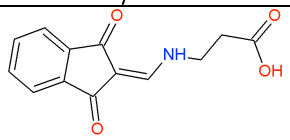
Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
Dasatinib		488.00	n/a	100	0.024±0.001	n/a
1		182.24	Binder SNAP 80 ms	59	82 ±20	n/a
2		170.23	Binder SNAP 80 ms	16	110±26	800±400
3		193.25	Binder SNAP 80 ms	12	n/a	900±400
4		242.26	Binder SNAP 80 ms	32	135±40	n/a
5		181.21	Binder SNAP 2 ms	22	700±500	n/a
6		260.25	Binder SNAP 80 ms	14	200±150	n/a

Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
7		186.17	Binder SNAP 80 ms	18	600±400	650±65
8		188.23	Binder SNAP 80 ms	14	7000±1200	1100±700
9		205.26	Binder SNAP 80 ms	7.50	n/a	n/a
10		202.26	Binder SNAP 80 ms	6	90±54	n/a
11		247.31	Binder SNAP 2 ms	9	680±300	n/a
12		165.24	Binder SNAP 80 ms	9	5500±800	n/a
13		249.31	Binder SNAP 80 ms	9	250±200	n/a
14		217.27	Binder SNAP 2 ms	12	200±150	n/a
15		204.25	Binder SNAP 80 ms	10	400±250	n/a

Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
16		182.2	Binder SNAP 80 ms	42	160±70	n/a
17		195.22	Binder SNAP 80 ms	45	120±50	n/a
18		273.33	Binder SNAP 2 ms	12	440±200	n/a
19		212.27	Binder SNAP 80 ms	18	240±100	n/a
20		246.24	Non binder SNAP 80 ms	21	500±250	n/a
21		201.23	Binder SNAP 80 ms	19	500±200	n/a
22		177.21	Binder SNAP 80 ms	14	434±475	n/a
23		239.33	Binder SNAP 80 ms	23	572±604	n/a

Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
24		188.23	*In silico*	80	155±60	n/a
25		203.23	*In silico*	n/a	n/a	n/a
26		264.33	*In silico*	n/a	25±5	n/a
27		296.34	*In silico*	n/a	4.5 ± 0.5	n/a
28		379.00	Publication	n/a	1.5	n/a
29		427.00	Publication	n/a	2.1	n/a
30		237.20	Binder SNAP 80 ms	2	n/a	950±400

Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
31		218.25	Binder SNAP 80 ms	-5	n/a	760±600
32		134.14	Binder SNAP 80 ms	5	n/a	900±300
33		208.21	Binder SNAP 80 ms	0	n/a	1000±400
34		207.27	Binder SNAP 80 ms	1	n/a	2000±1400
35		193.25	Binder SNAP 80 ms	0	n/a	970±400
36		179.22	Binder SNAP 80 ms	-5	n/a	1400±1000
37		264.28	Binder SNAP 80 ms	1	n/a	1700±600
38		176.22	Binder SNAP 80 ms	-1.5	n/a	2000±1500

Compound	Structure	Mw	Origin[a]	Bioassay at 500 µM PIN (%)[b]	IC_{50} (µM)	K_D[c] (µM)
39		181.19	Binder Biot 2 ms	0	n/a	900±400
40		191.24	Binder Biot 2 ms	n/a	n/a	1200±300
41		215.26	Binder Biot 2 ms	n/a	n/a	1300±600
42		161.16	Binder Biot 2 ms	n/a	n/a	2000±900
43		183.23	Binder Biot 2 ms	3	n/a	2700±1700
44		245.23	Binder Biot 2 ms	n/a	n/a	460±60

[a] Origin of the compound: fragments from the TINS screen made on the SNAP fusion KD at either 80 ms or 2 ms (Binder SNAP 80 ms or 2ms), Non binders from the same screen (Non binder SNAP 80 ms), compounds selected or optimized during the *in silico* screen (*In silico*), hits selected from the TINS screen made on the biotinylated KD (Binder Biot 2 ms)

[b] Percentage of inhibition (P.I.N) of the compound at a concentration of 500 µM in the biochemical assay. Above 6% of inhibition, compounds are considered as inhibitory.

[c] Binding affinity determined with Surface Plasmon Resonance assay (SPR).

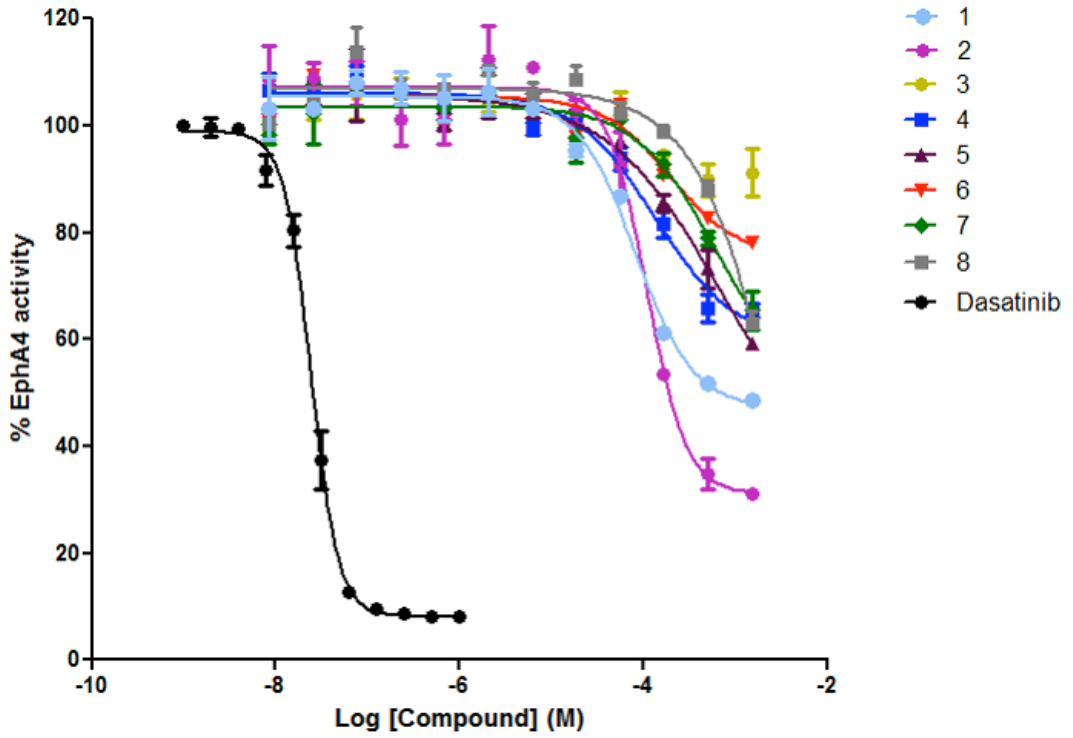
Appendix B. Enzyme Inhibition Data for Selected TINS screening hits.

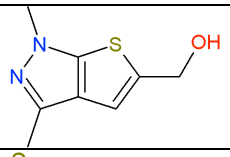
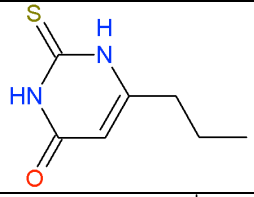
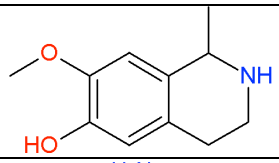
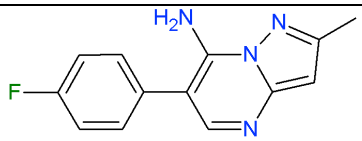
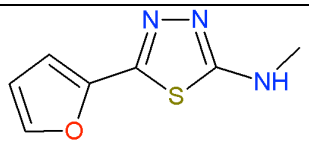
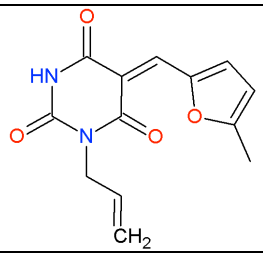
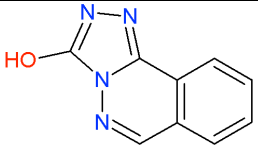
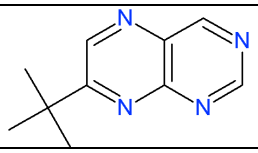
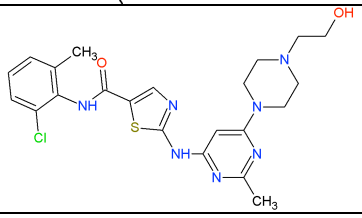
Enzyme inhibition curves for selected hits from the TINS screen. All data was generated by Wim Zoutman in the laboratory of Cornelis Tensen at the LUMC. As a reference, the dasatinib dose-response curve is shown in black. Each curve represents the mean ± S.E.M of experiments performed in triplicate. Curve fitting was performed using GraphPad Prism version 5.0 using a sigmoidal dose response curve model with variable slope. Inhibitory constants, upper and lower asymptote values and Hill coefficients for each fragment are provided in the table.

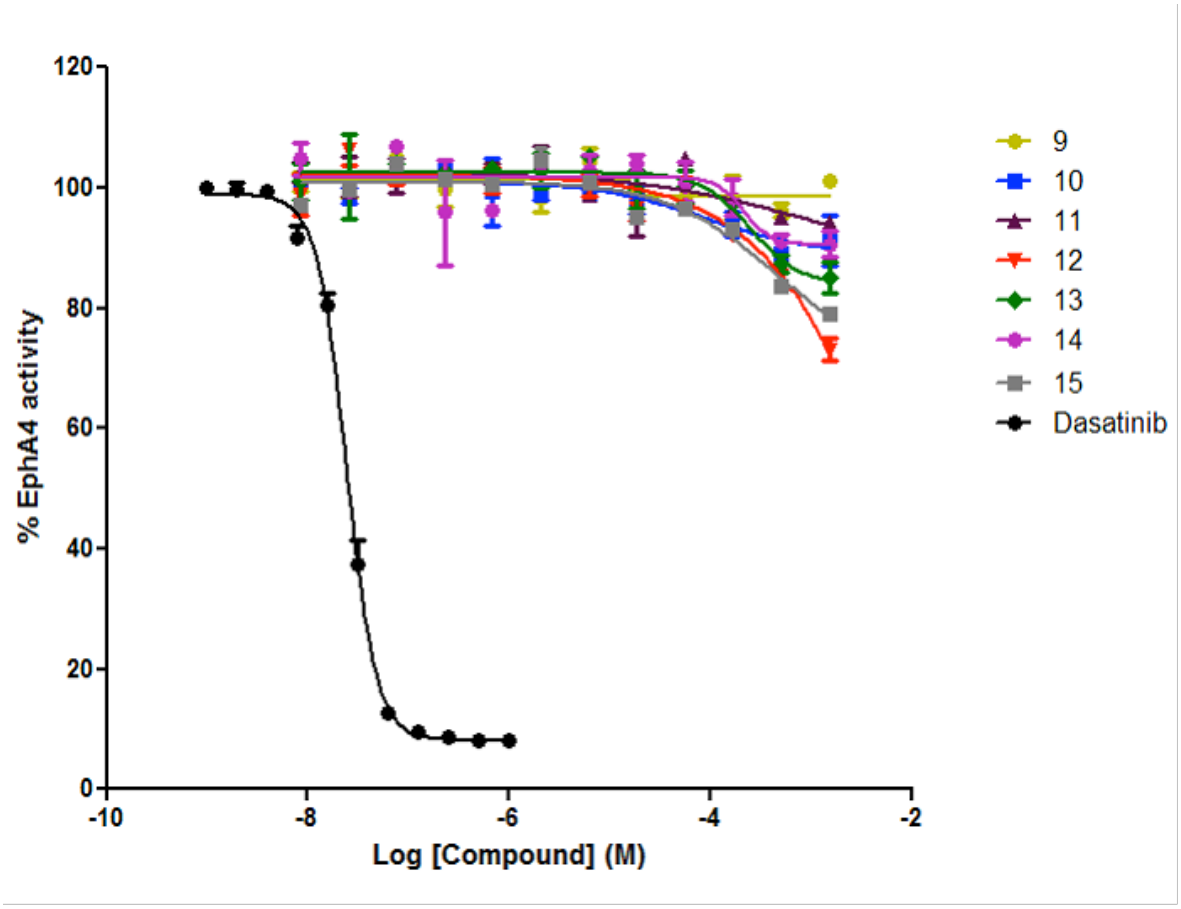
In the program Prism, the curves are fitted to the equation:

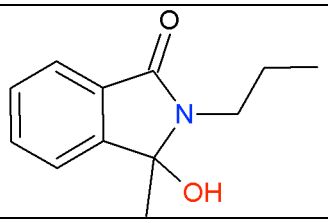
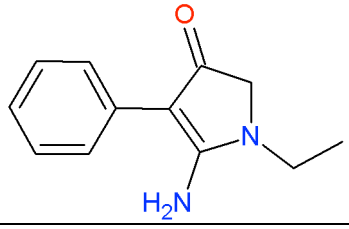
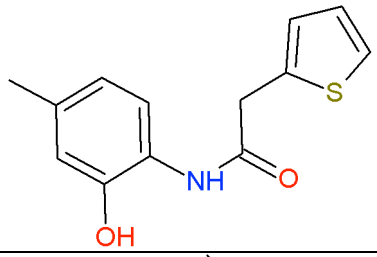
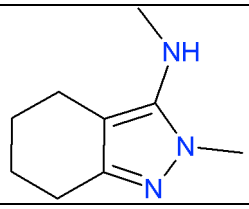
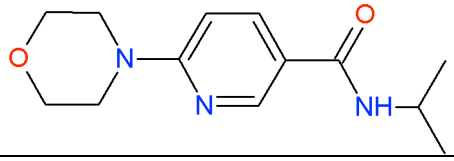
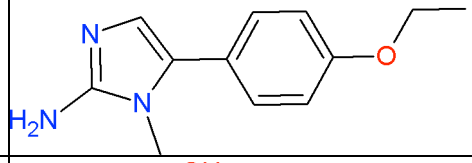
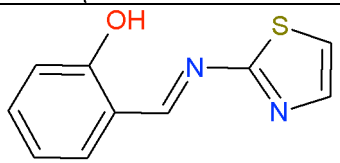
$$y = \text{bottom} + [(\text{top-bottom}) / (1+10^{(\text{Log IC50}-x) - \text{Hill slope}})]$$

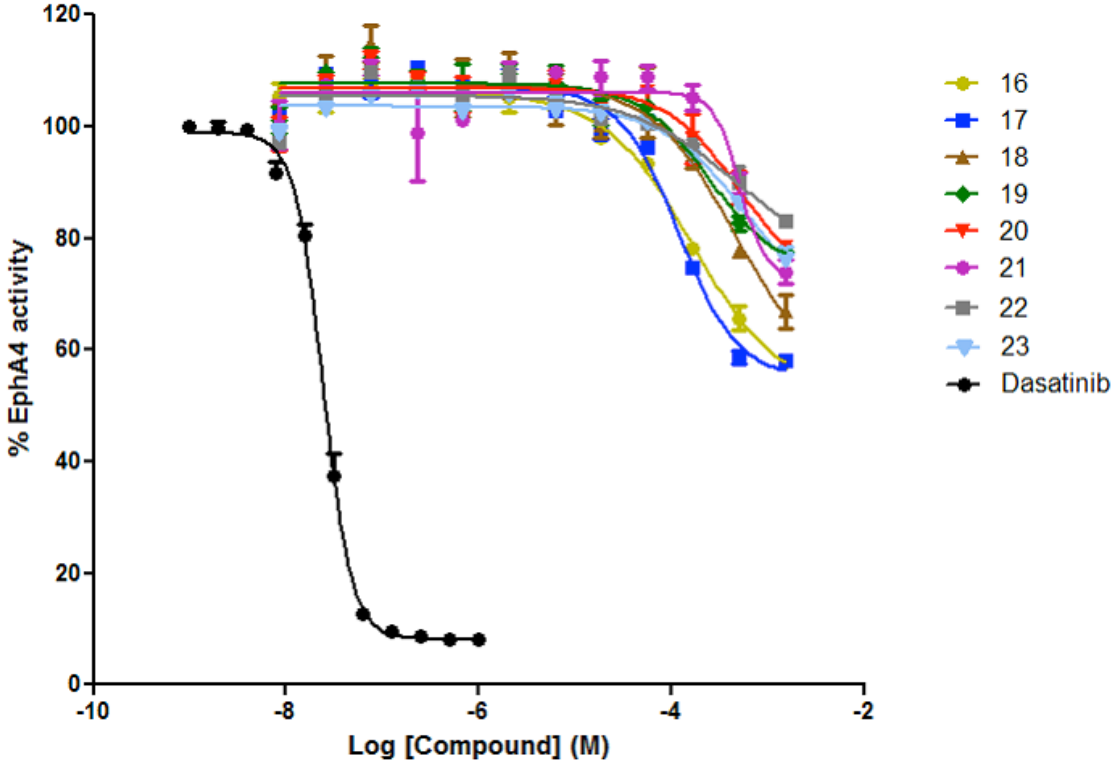
where top is the upper asymptote values and bottom is the lower asymptote value.

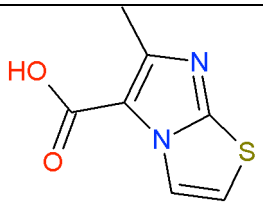
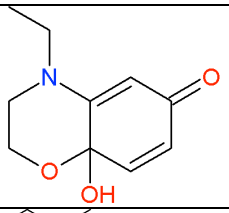
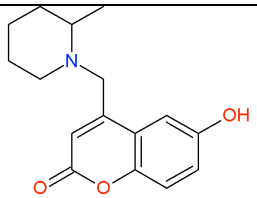
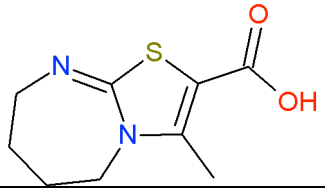
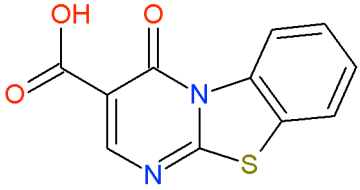
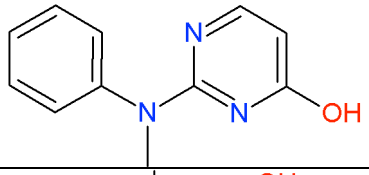
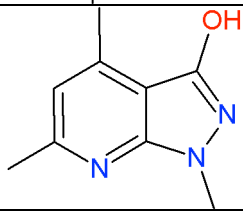
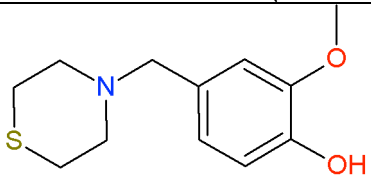


Compound	Structure	IC_{50} (μM)	Y value at the top plateau of the curve	Y value at the bottom plateau of the curve	Hill coefficient
1		82 ±20	105.3±0.9	46.9±2.3	-1.40±0.18
2		110±26	107.0±1.2	30.9±2.9	-2.04±0.33
3		n/a			
4		135±40	106±1	59±4.0	-0.99±0.18
5		700±500	105±1	30±27	-0.70±0.18
6		200±150	105±1.5	75.5±7.0	-1.19±0.61
7		600±400	103.5±1.0	50 ±18	-1.02±0.36
8		7000±1200	106.9±1.2	69±29	-0.87±0.50
Dasatinib		0.024±0.001	98.8±0.9	8.1±0.8	-2.91±0.17


120
100
80
60
40
20
0
% EphA4 activity
-10
-8
-6
-4
-2
Log [Compound] (M)
9
10
11
12
13
14
15
Dasatinib

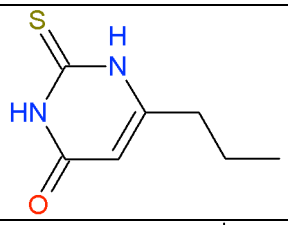
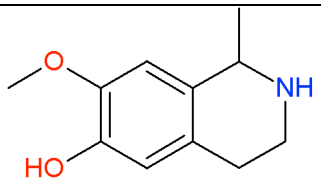
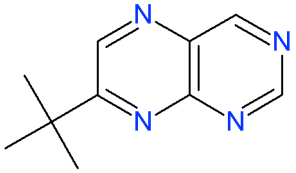
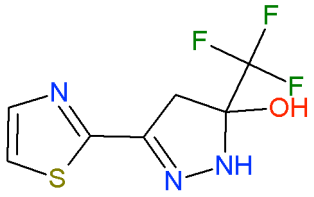
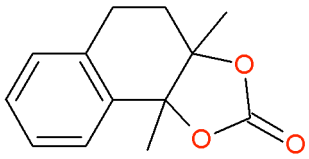
Compound	Structure	IC_{50} (μM)	Y value at the top plateau of the curve	Y value at the bottom plateau of the curve	Hill coefficient
9		n/a			
10		90±54	100.9±0.8	88.7±3.6	-0.80±0.49
11		680±300	102.3±1.2	88±27	-0.54±0.74
12		5500±800	102.0±0.9	2±200	-0.65±0.31
13		250±200	102.4±0.6	83±3	-1.70±0.78
14		200±150	101.6±1.0	90±3	-3.28±1.01
15		400±250	101.0±0.8	70±12	-0.81±0.33

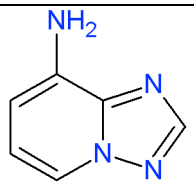
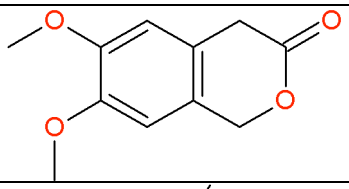
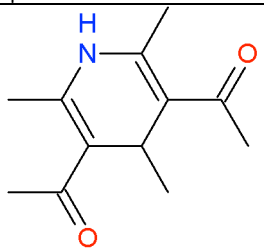
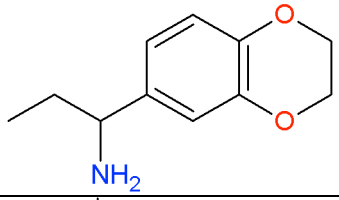
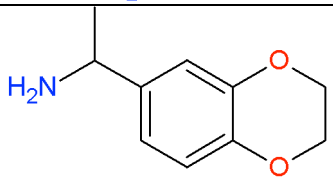
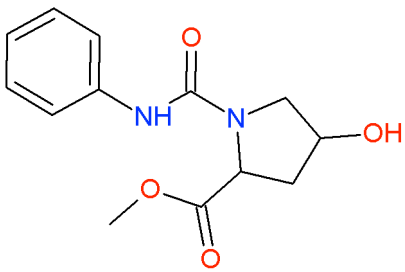
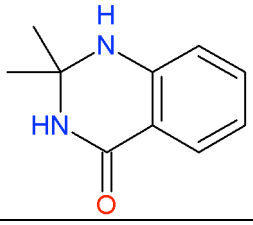

120
100
80
60
40
20
0
% EphA4 activity
-10
-8
-6
-4
-2
Log [Compound] (M)
16
17
18
19
20
21
22
23
Dasatinib

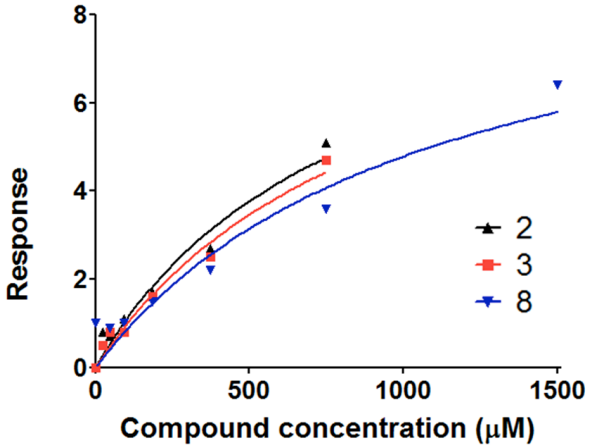
Compound	Structure	IC_{50} (μM)	Y value at the top plateau of the curve	Y value at the bottom plateau of the curve	Hill coefficient
16		160±70	106.0±0.6	51±3	-0.95±0.11
17		120±50	107±1	55±3	-1.42±0.25
18		440±200	107 ±1.2	54±16	-1.01±0.38
19		240±100	107.6±0.9	73±5	-1.21±0.39
20		500±250	107±2	70±13	-1.16±0.56
21		500±200	106.0±1.1	73±6	-3.32±4.23
22		434±475	105.5±1.1	65±16	-0.82±0.47
23		572±604	103.5±0.6	87±12	-1.01±0.33

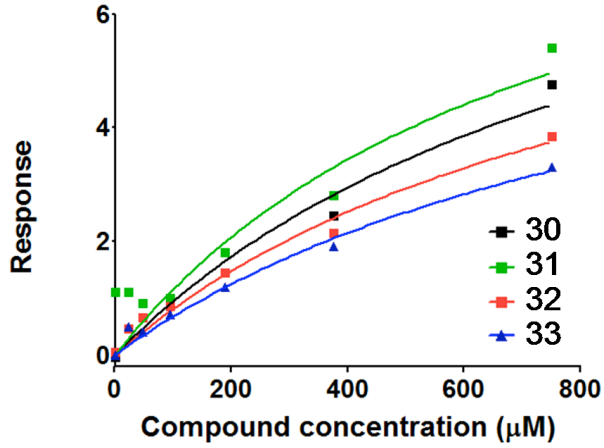
Appendix C: SPR analysis of fragment binding to EphA4 KD immobilized via the 6His tag.

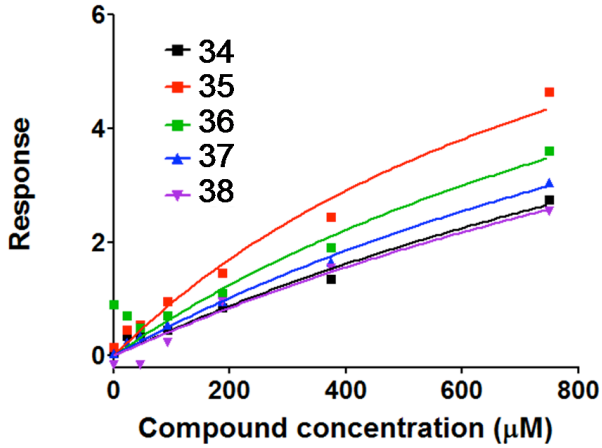
In Chapter 6, SPR experiments were conducted on the 6His EphA4-KD. After elimination of ill-behaved compounds, 73 out of 134 compounds were considered as hits in this SPR assay. Twelve of them presented a clear concentration dependent response curve. The representative saturation curve of each compound and an estimation of the equilibrium dissociation constant are summarized in the table below. The response signal is plotted against the concentration of the compound.

Compound	Structure	K_D (μM)
2		800±400
3		900±400
8		1100±700
30		950±400
31		760±600

Compound	Structure	K_D (μM)
32		900±300
33		1000±400
34		2000±1400
35		970±400
36		1400±1000
37		1700±600
38		2000±1500

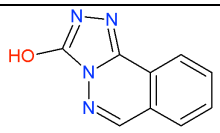
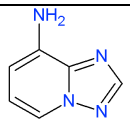
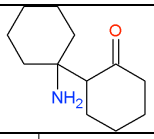
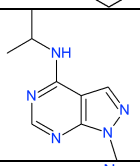
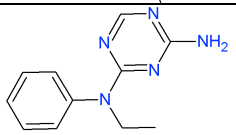
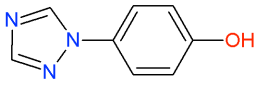
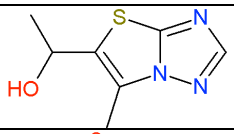
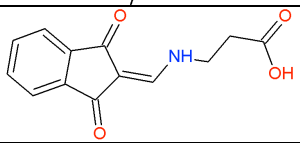

Response
Compound concentration (μM)
2
3
8

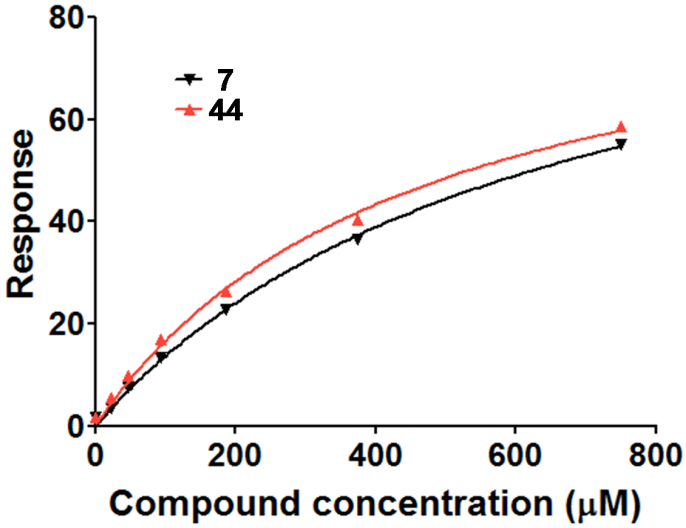

Response
Compound concentration (μM)
30
31
32
33

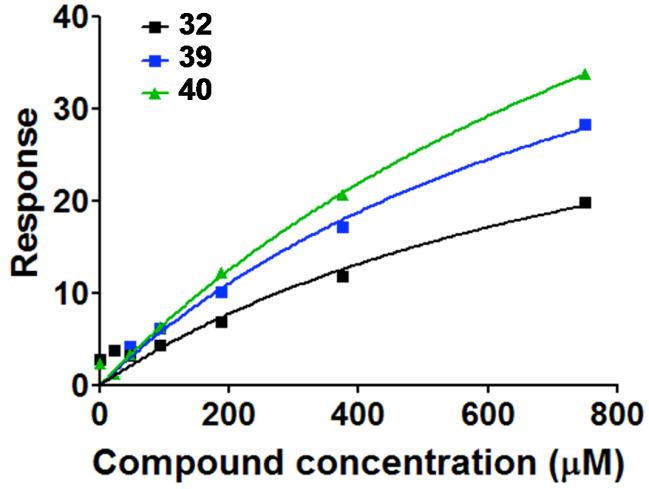

Response
Compound concentration (μM)
34
35
36
37
38

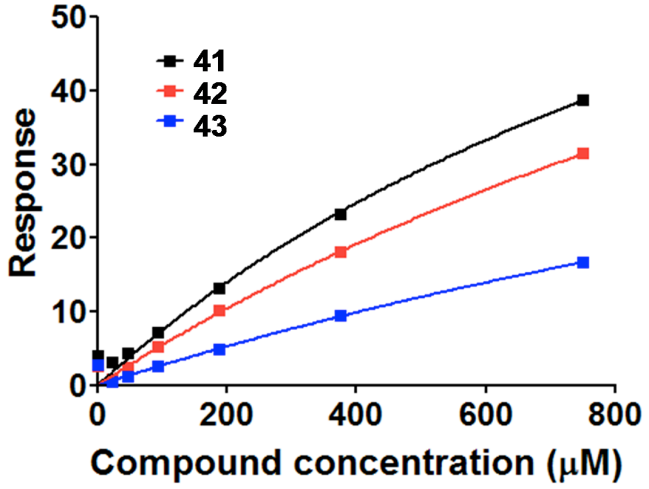
Appendix D: SPR analysis of fragment binding to EphA4 KD immobilized via NeutrAvidin.

In Chapter 6, SPR experiments were conducted on EphA4 KD immobilized via NeutrAvidin. A total of 62 compounds were selected as hits in the TINS experiment on the biotinylated EphA4 (Chapter 3) and assayed for binding by SPR. Out of these 62 compounds, 51 were positive for binding and reproducible concentration dependent curve could be obtained for 8 of them. A representative saturation curve of each compound and an estimation of the equilibrium dissociation constant are summarized in the table below. The response signal is plotted against the concentration of the compound.

Compound	Structure	K_D (μM)
7		650±65
32		900±300
39		900±400
40		1200±300
41		1300±600
42		2000±900
43		2700±1700
44		460±60

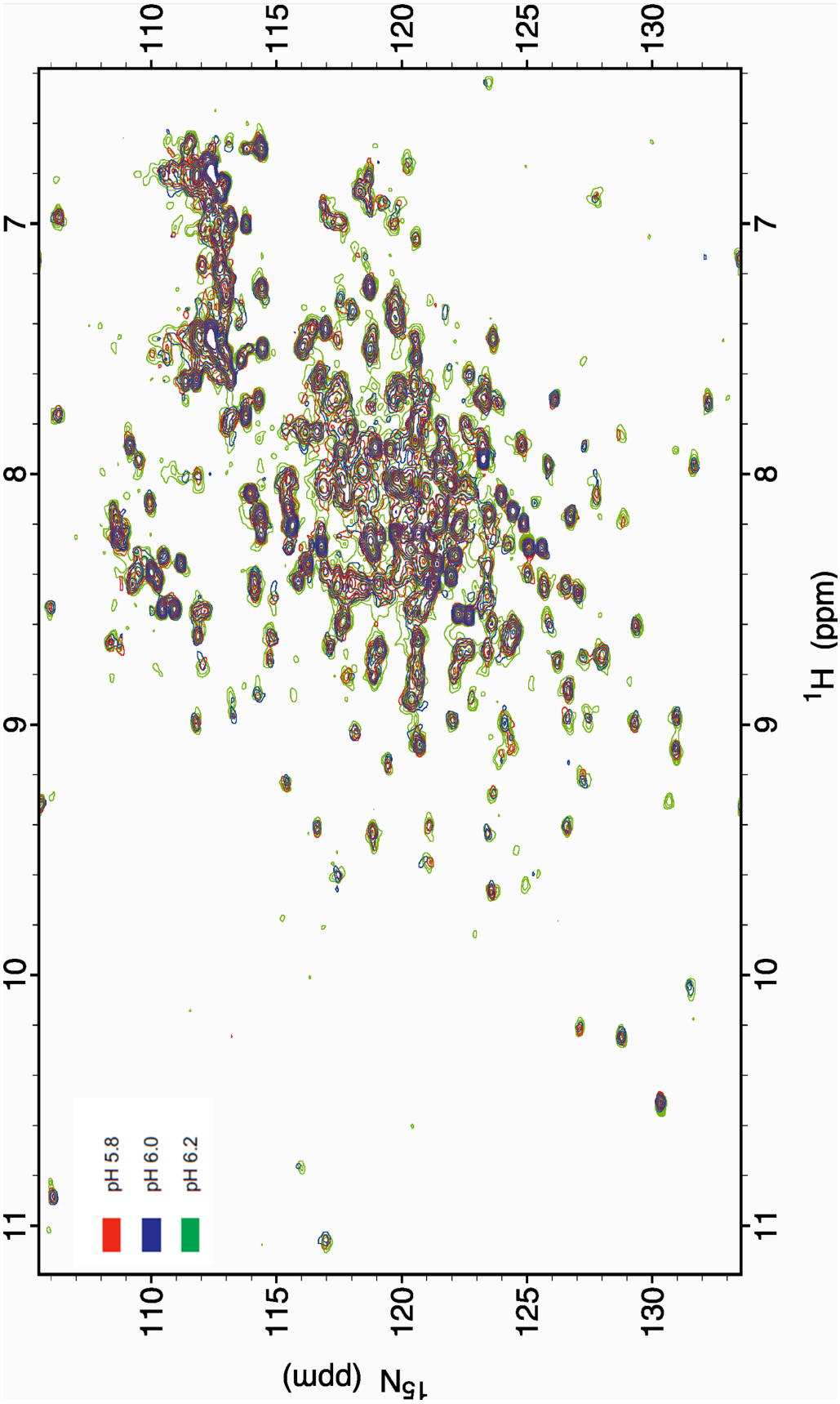

7
44
Response
Compound concentration (μM)
0
20
40
60
80
0
200
400
600
800


32
39
40
Response
Compound concentration (μM)
0
10
20
30
40
0
200
400
600
800


41
42
43
Response
Compound concentration (μM)
0
10
20
30
40
50
0
200
400
600
800

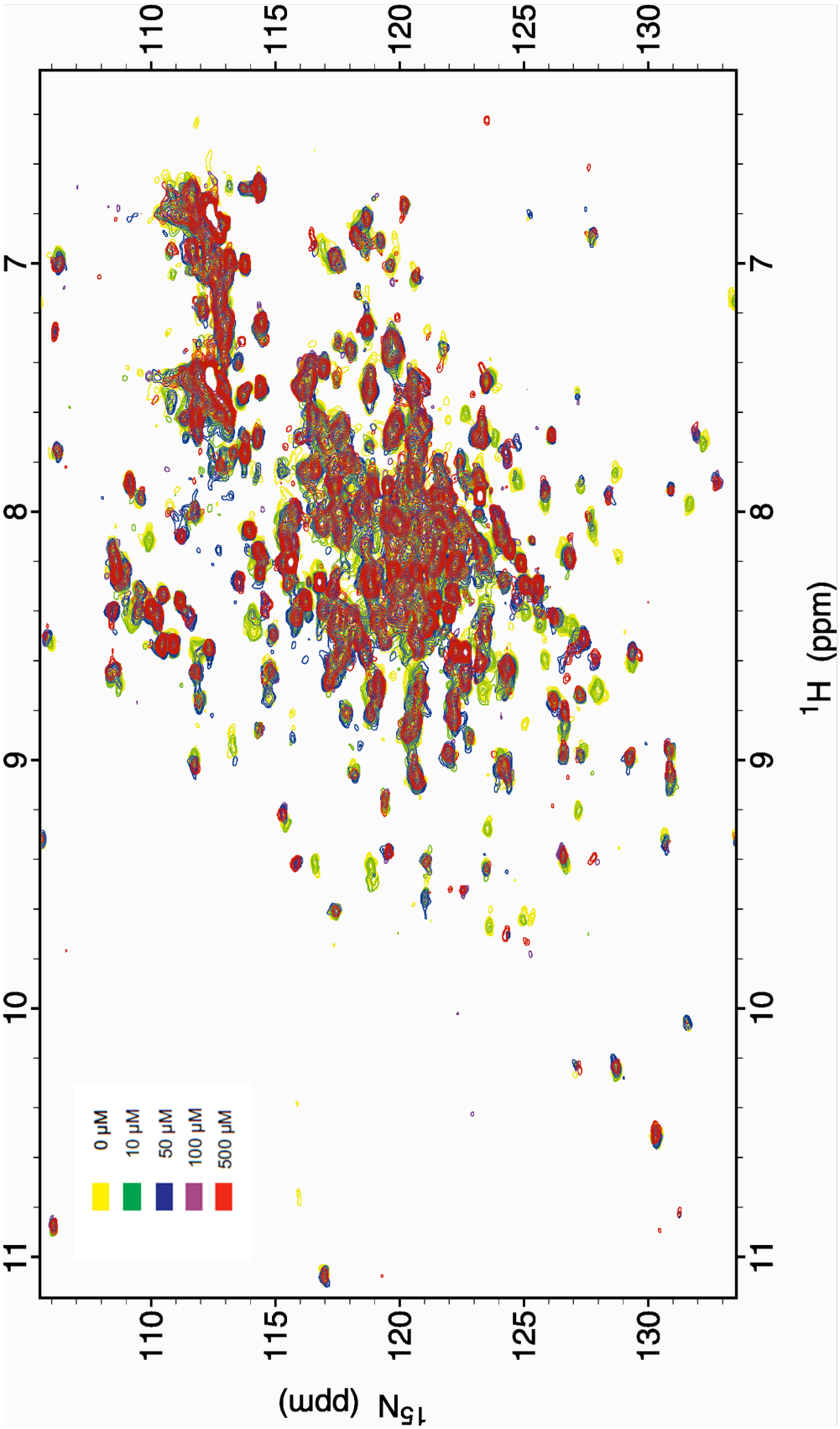
Appendix E: pH titration of ^{15}N EphA4 KD.

The overlaid [^{15}N,^{1}H]-HSQC spectra of 100 uM EphA4 KD is recorded at pH 5.8, 6.0 and 6.2.


pH 5.8
pH 6.0
pH 6.2
^{15}N (ppm)
^{1}H (ppm)

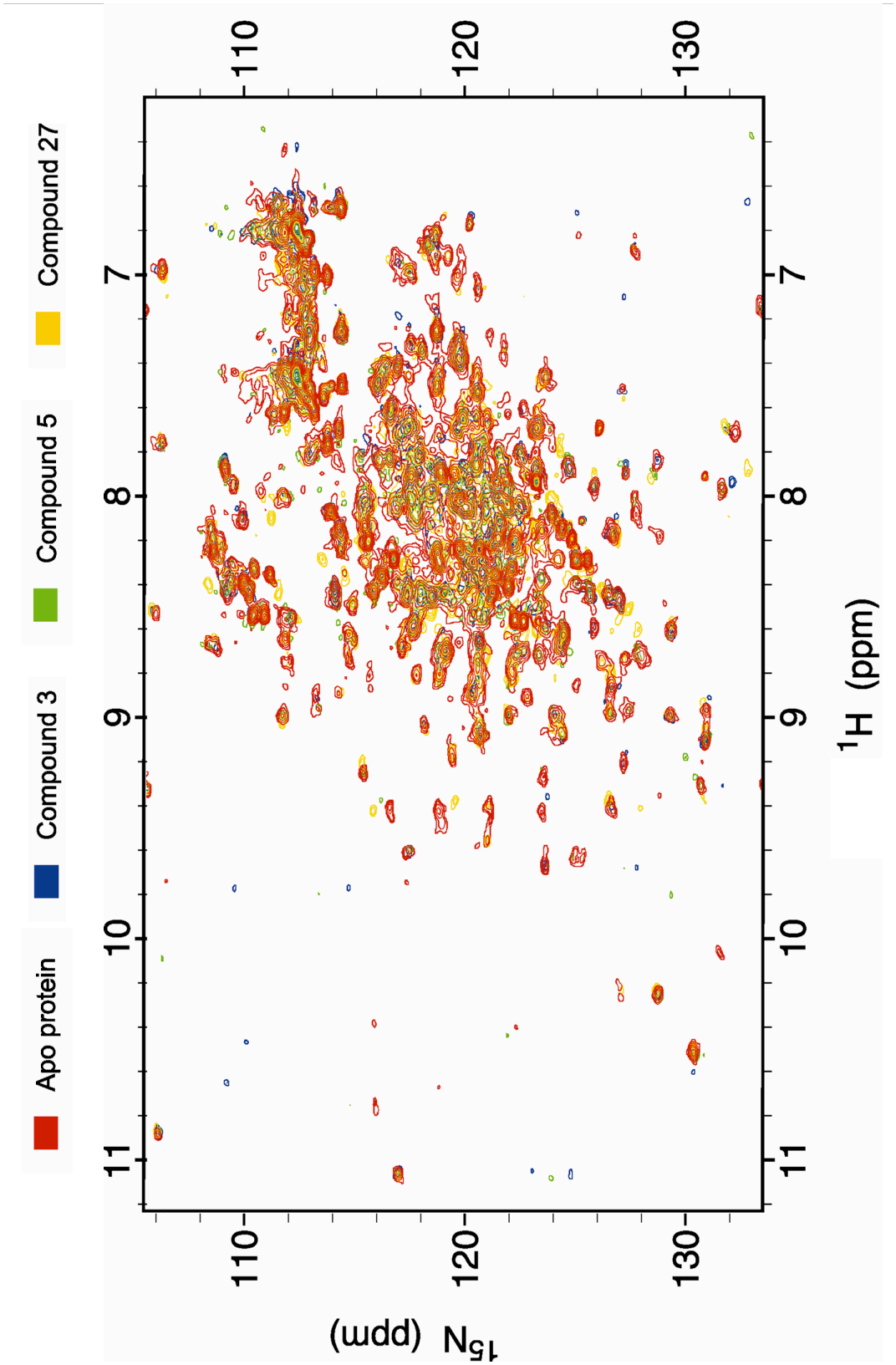
Appendix F: NMR analysis of compound 27 binding to ^{15}N EphA4 KD.

Overlay of [^{15}N,^{1}H]-HSQC spectra of 100 μM of ^{15}N EphA4-KD with increasing concentration of compound **27** as indicated.


0 µM
10 µM
50 µM
100 µM
500 µM
^{15}N (ppm)
^{1}H (ppm)

Appendix G: NMR analysis of ligand binding to KD EphA4

A region of the overlaid [^{15}N,^{1}H]-HSQC spectra of 100 μM ^{15}N-KD is presented. Overlay of the spectra of: apo –protein (red), EphA4 KD upon titration of 2 mM compound **3** (blue), EphA4 KD upon titration of 2 mM of compound **5** (green) and EphA4 KD upon titration of 100 μM of compound **27** (yellow).


Apo protein
Compound 3
Compound 5
Compound 27
110
120
130
7
8
9
10
11
^{15}N (ppm)
^{1}H (ppm)

References

1. J. F. Borzelleca, *Toxicol. Sci.*, 2000, **53**, 2-4.
2. F. W. Sertürner, *Gilbert's Annalen d. Physik*, 1817, **25**, 56.
3. W. Sneader, *BMJ*, 2000, **321**, 1591-1594.
4. A. Fleming, *Br. J. Exp. Pathol.*, 1940, **10**, 226.
5. H. C. Hinshaw, M. M. Pyle and W. H. Feldman, *Am J Med*, 1947, **2**, 429-435.
6. A. Giner-Sorolla, *Trends Pharmacol Sci*, 1988, **9**, 437-438.
7. J. D. Watson and F. H. Crick, *Cold Spring Harb Symp Quant Biol*, 1953, **18**, 123-131.
8. M. Branca, *Bio-IT World April 2003*, 2003.
9. S. N. Cohen, A. C. Y. Chang, H. W. Boyer and R. B. Helling, *Proc Natl Acad Sci U S A*, 1973, **70**, 3240-3244.
10. D. A. Jackson, P. Berg and R. H. Symons, *Proc Natl Acad Sci U S A*, 1972, **69**, 2904-&.
11. M. S. Gottlieb, R. Schroff, H. M. Schanker, J. D. Weisman, P. T. Fan, R. A. Wolf and A. Saxon, *N Engl J Med*, 1981, **305**, 1425-1431.
12. J. A. Levy, *Microbiol Rev*, 1993, **57**, 183-289.
13. J. M. Bartlett and D. Stirling, *Methods Mol Biol*, 2003, **226**, 3-6.
14. S. Y. Pan, S. Pan, Z. L. Yu, D. L. Ma, S. B. Chen, W. F. Fong, Y. F. Han and K. M. Ko, *J Pharm Pharm Sci*, 2010, **13**, 450-471.
15. M. A. Lindsay, *Nat Rev Drug Discov*, 2003, **2**, 831-838.
16. M. E. Nuttall, *Cells Tissues Organs*, 2001, **169**, 265-271.
17. S. Fox, S. Farr-Jones, L. Sopchak, A. Boggs, H. W. Nicely, R. Khoury and M. Biros, *J Biomol Screen*, 2006, **11**, 864-869.
18. G. M. Rishton, *Drug Discov Today*, 1997, **2**, 382-384.
19. S. L. McGovern, E. Caselli, N. Grigorieff and B. K. Shoichet, *J Med Chem*, 2002, **45**, 1712-1722.
20. I. G. Popa-Burke, O. Issakova, J. D. Arroway, P. Bernasconi, M. Chen, L. Coudurier, S. Galasinski, A. P. Jadhav, W. P. Janzen, D. Lagasca, D. Liu, R. S. Lewis, R. P. Mohney, N. Sepetov, D. A. Sparkman and C. N. Hodge, *Anal Chem*, 2004, **76**, 7278-7287.
21. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv Drug Deliv Rev*, 2001, **46**, 3-26.
22. T. I. Oprea, A. M. Davis, S. J. Teague and P. D. Leeson, *J Chem Inf Comput Sci*, 2001, **41**, 1308-1315.
23. S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angew Chem Int Ed Engl*, 1999, **38**, 3743-3748.
24. W. P. Jencks, *Proc Natl Acad Sci U S A*, 1981, **78**, 4046-4050.
25. N. M. Green, *Adv Protein Chem*, 1975, **29**, 85-133.
26. C. E. Nakamura and R. H. Abeles, *Biochemistry*, 1985, **24**, 1364-1376.
27. S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik, *Science*, 1996, **274**, 1531-1534.

28. M. Congreve, R. Carr, C. Murray and H. Jhoti, *Drug Discov Today*, 2003, **8**, 876-877.
29. R. A. Carr, M. Congreve, C. W. Murray and D. C. Rees, *Drug Discov Today*, 2005, **10**, 987-992.
30. R. S. Bohacek, C. McMartin and W. C. Guida, *Med Res Rev*, 1996, **16**, 3-50.
31. T. Fink, H. Bruggesser and J. L. Reymond, *Angewandte Chemie*, 2005, **44**, 1504-1508.
32. P. J. Hajduk and J. Greer, *Nature reviews. Drug discovery*, 2007, **6**, 211-219.
33. D. A. Erlanson, *Top Curr Chem*, 2011.
34. P. G. Wyatt, A. J. Woodhead, V. Berdini, J. A. Boulstridge, M. G. Carr, D. M. Cross, D. J. Davis, L. A. Devine, T. R. Early, R. E. Feltell, E. J. Lewis, R. L. McMenamin, E. F. Navarro, M. A. O'Brien, M. O'Reilly, M. Reule, G. Saxty, L. C. Seavers, D. M. Smith, M. S. Squires, G. Trewartha, M. T. Walker and A. J. Woolford, *J Med Chem*, 2008, **51**, 4986-4999.
35. S. Howard, V. Berdini, J. A. Boulstridge, M. G. Carr, D. M. Cross, J. Curry, L. A. Devine, T. R. Early, L. Fazal, A. L. Gill, M. Heathcote, S. Maman, J. E. Matthews, R. L. McMenamin, E. F. Navarro, M. A. O'Brien, M. O'Reilly, D. C. Rees, M. Reule, D. Tisi, G. Williams, M. Vinkovic and P. G. Wyatt, *J Med Chem*, 2009, **52**, 379-388.
36. G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. A. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, K. D'Andrea, A. Koehler, M. Stumm, P. S. Lin, R. J. Lee, J. Grippo, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. B. Chapman, K. T. Flaherty, X. Xu, K. L. Nathanson and K. Nolop, *Nature*, 2010, **467**, 596-599.
37. K. T. Flaherty, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. J. O'Dwyer, R. J. Lee, J. F. Grippo, K. Nolop and P. B. Chapman, *N Engl J Med*, 2010, **363**, 809-819.
38. R. Macarron, *Drug Discov Today*, 2006, **11**, 277-279.
39. S. Vanwetswinkel, R. J. Heetebrij, J. van Duynhoven, J. G. Hollander, D. V. Filippov, P. J. Hajduk and G. Siegal, *Chem Biol*, 2005, **12**, 207-216.
40. T. Marquardsen, M. Hofmann, J. G. Hollander, C. M. Loch, S. R. Kiihne, F. Engelke and G. Siegal, *J Magn Reson*, 2006, **182**, 55-65.
41. C. Ortutay, J. Valiaho, K. Stenberg and M. Vihinen, *Hum Mutat*, 2005, **25**, 435-442.
42. J. Schlessinger, *Cell*, 2000, **103**, 211-225.
43. E. Zwick, J. Bange and A. Ullrich, *Endocr Relat Cancer*, 2001, **8**, 161-173.
44. K. S. Kolibaba and B. J. Druker, *Biochim Biophys Acta*, 1997, **1333**, F217-248.
45. B. J. Druker, M. Talpaz, D. J. Resta, B. Peng, E. Buchdunger, J. M. Ford, N. B. Lydon, H. Kantarjian, R. Capdeville, S. Ohno-Jones and C. L. Sawyers, *N Engl J Med*, 2001, **344**, 1031-1037.
46. H. Commander, G. Whiteside and C. Perry, *Drugs*, 2011, **71**, 1355-1365.
47. M. W. Karaman, S. Herrgard, D. K. Treiber, P. Gallant, C. E. Atteridge, B. T. Campbell, K. W. Chan, P. Ciceri, M. I. Davis, P. T. Edeen, R. Faraoni, M. Floyd, J. P. Hunt, D. J. Lockhart, Z. V. Milanov, M. J. Morrison, G. Pallares, H. K. Patel,

S. Pritchard, L. M. Wodicka and P. P. Zarrinkar, *Nat Biotechnol*, 2008, **26**, 127-132.
48. O. Fedorov, S. Muller and S. Knapp, *Nat Chem Biol*, 2010, **6**, 166-169.
49. P. van der Geer, T. Hunter and R. A. Lindberg, *Annu Rev Cell Biol*, 1994, **10**, 251-337.
50. N. Cheng, D. M. Brantley and J. Chen, *Cytokine & growth factor reviews*, 2002, **13**, 75-85.
51. E. B. Pasquale, *Cell*, 2008, **133**, 38-52.
52. E. N. Committee, *Unified nomenclature for Eph family receptors and their ligands, the ephrins* 0092-8674 (Print)
0092-8674 (Linking), 1997.
53. E. B. Pasquale, *Nat Rev Mol Cell Biol*, 2005, **6**, 462-475.
54. A. D. Bergemann, L. Zhang, M. K. Chiang, R. Brambilla, R. Klein and J. G. Flanagan, *Oncogene*, 1998, **16**, 471-480.
55. N. W. Gale, A. Flenniken, D. C. Compton, N. Jenkins, N. G. Copeland, D. J. Gilbert, S. Davis, D. G. Wilkinson and G. D. Yancopoulos, *Oncogene*, 1996, **13**, 1343-1352.
56. N. W. Gale, S. J. Holland, D. M. Valenzuela, A. Flenniken, L. Pan, T. E. Ryan, M. Henkemeyer, K. Strebhardt, H. Hirai, D. G. Wilkinson, T. Pawson, S. Davis and G. D. Yancopoulos, *Neuron*, 1996, **17**, 9-19.
57. J. Gong, A. Morishita, K. Kurokohchi, J. Tani, K. Kato, H. Miyoshi, H. Inoue, M. Kobayashi, S. Liu, M. Murota, A. Muramatsu, K. Izuishi, Y. Suzuki, H. Yoshida, N. Uchida, K. Deguchi, H. Iwama, I. Ishimaru and T. Masaki, *Int J Oncol*, 2010, **36**, 101-106.
58. R. van Doorn, R. Dijkman, M. H. Vermeer, J. J. Out-Luiting, E. M. van der Raaij-Helmer, R. Willemze and C. P. Tensen, *Cancer Res*, 2004, **64**, 5578-5586.
59. S. Ashida, H. Nakagawa, T. Katagiri, M. Furihata, M. Iiizumi, Y. Anazawa, T. Tsunoda, R. Takata, K. Kasahara, T. Miki, T. Fujioka, T. Shuin and Y. Nakamura, *Cancer Res*, 2004, **64**, 5963-5972.
60. Y. Goldshmit, M. P. Galea, G. Wise, P. F. Bartlett and A. M. Turnley, *J Neurosci*, 2004, **24**, 10064-10073.
61. Y. Goldshmit, M. D. Spanevello, S. Tajouri, L. Li, F. Rogers, M. Pearse, M. Galea, P. F. Bartlett, A. W. Boyd and A. M. Turnley, *PloS one*, 2011, **6**, e24636.
62. G. Siegal, E. Ab and J. Schultz, *Drug Discov Today*, 2007, **12**, 1032-1039.
63. C. L. Verlinde, E. Fan, S. Shibata, Z. Zhang, Z. Sun, W. Deng, J. Ross, J. Kim, L. Xiao, T. L. Arakaki, J. Bosch, J. M. Caruthers, E. T. Larson, I. Letrong, A. Napuli, A. Kelly, N. Mueller, F. Zucker, W. C. Van Voorhis, F. S. Buckner, E. A. Merritt and W. G. Hol, *Curr Top Med Chem*, 2009, **9**, 1678-1687.
64. K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho and K. Wuthrich, *Proc Natl Acad Sci U S A*, 1998, **95**, 14147-14151.
65. P. J. Hajduk, D. J. Augeri, J. Mack, R. Mendoza, J. G. Yang, S. F. Betz and S. W. Fesik, *J Am Chem Soc*, 2000, **122**, 7898-7904.
66. D. F. Wyss, Y. S. Wang, H. L. Eaton, C. Strickland, J. H. Voigt, Z. Zhu and A. W. Stamford, *Top Curr Chem*, 2011, In press.
67. D. J. Liu, Y. S. Wang, J. J. Gesell, E. Wilson, B. M. Beyer and D. F. Wyss, *J Biomol NMR*, 2004, **29**, 425-426.

68. H. L. Eaton and D. F. Wyss, *Methods Enzymol*, 2011, **493**, 447-468.
69. M. Krajewski, U. Rothweiler, L. D'Silva, S. Majumdar, C. Klein and T. A. Holak, *J Med Chem*, 2007, **50**, 4382-4387.
70. U. Rothweiler, A. Czarna, L. Weber, G. M. Popowicz, K. Brongel, K. Kowalska, M. Orth, O. Stemmann and T. A. Holak, *J Med Chem*, 2008, **51**, 5035-5042.
71. M. Bista, K. Kowalska, W. Janczyk, A. Domling and T. A. Holak, *J Am Chem Soc*, 2009, **131**, 7500-7501.
72. D. Grandy, J. Shan, X. Zhang, S. Rao, S. Akunuru, H. Li, Y. Zhang, I. Alpatov, X. A. Zhang, R. A. Lang, D. L. Shi and J. J. Zheng, *J Biol Chem*, 2009, **284**, 16256-16263.
73. J. Shan and J. J. Zheng, *J Comput Aided Mol Des*, 2009, **23**, 37-47.
74. V. Fruh, Y. Zhou, D. Chen, C. Loch, E. Ab, Y. N. Grinkova, H. Verheij, S. G. Sligar, J. H. Bushweller and G. Siegal, *Chem Biol*, 2010, **17**, 881-891.
75. M. Congreve, R. L. Rich, D. G. Myszka, F. Figaroa, G. Siegal and F. H. Marshall, *Methods Enzymol*, 2011, **493**, 115-136.
76. W. Jahnke, L. B. Perez, C. G. Paris, A. Strauss, G. Fendrich and C. M. Nalin, *J Am Chem Soc*, 2000, **122**, 7394-7395.
77. E. Balogh, D. Wu, G. Zhou and M. Gochin, *J Am Chem Soc*, 2009, **131**, 2821-2823.
78. T. Saio, K. Ogura, K. Shimizu, M. Yokochi, T. R. Burke, Jr. and F. Inagaki, *J Biomol NMR*, 2011, **51**, 395-408.
79. M. Leone, E. Barile, J. Vazquez, A. Mei, D. Guiney, R. Dahl and M. Pellecchia, *Chem Biol Drug Des*, 2010, **76**, 10-16.
80. M. Mayer and B. Meyer, *Angew Chem Int Ed Engl*, 1999, **38**, 1784-1788.
81. M. Mayer and B. Meyer, *J Am Chem Soc*, 2001, **123**, 6108-6117.
82. G. Timpano, G. Tabarani, M. Anderluh, D. Invernizzi, F. Vasile, D. Potenza, P. M. Nieto, J. Rojo, F. Fieschi and A. Bernardi, *Chembiochem*, 2008, **9**, 1921-1930.
83. M. Kobayashi, K. Retra, F. Figaroa, J. G. Hollander, E. Ab, R. J. Heetebrij, H. Irth and G. Siegal, *J Biomol Screen*, 2010, **15**, 978-989.
84. R. E. Hubbard and J. B. Murray, *Methods Enzymol*, 2011, **493**, 509-531.
85. K. E. Kover, P. Groves, J. Jimenez-Barbero and G. Batta, *J Am Chem Soc*, 2007, **129**, 11579-11582.
86. K. E. Kover, E. Weber, T. A. Martinek, E. Monostori and G. Batta, *Chembiochem*, 2010, **11**, 2182-2187.
87. A. Bhunia and S. Bhattacharjya, *Biopolymers*, 2011, **96**, 273-287.
88. J. P. Ribeiro, S. Andre, F. J. Canada, H. J. Gabius, A. P. Butera, R. J. Alves and J. Jimenez-Barbero, *ChemMedChem*, 2010, **5**, 415-419.
89. A. Maggioni, M. von Itzstein, J. Tiralongo and T. Haselhorst, *Chembiochem*, 2008, **9**, 2784-2786.
90. S. Bartoschek, T. Klabunde, E. Defossa, V. Dietrich, S. Stengelin, C. Griesinger, T. Carlomagno, I. Focken and K. U. Wendt, *Angew Chem Int Ed Engl*, 2010, **49**, 1426-1429.
91. B. Claasen, M. Axmann, R. Meinecke and B. Meyer, *J Am Chem Soc*, 2005, **127**, 916-919.
92. C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi and B. Stockman, *J Biomol NMR*, 2001, **21**, 349-359.

93. P. J. Hajduk, E. T. Olejniczak and S. W. Fesik, *J Am Chem Soc*, 1997, **119**, 12257-12261.
94. W. Jahnke, R. M. Grotzfeld, X. Pelle, A. Strauss, G. Fendrich, S. W. Cowan-Jacob, S. Cotesta, D. Fabbro, P. Furet, J. Mestan and A. L. Marzinzik, *J Am Chem Soc*, 2010, **132**, 7043-7048.
95. J. Hu, P. O. Eriksson and G. Kern, *Magn Reson Chem*, 2010, **48**, 909-911.
96. C. Ludwig, P. J. Michiels, X. Wu, K. L. Kavanagh, E. Pilka, A. Jansson, U. Oppermann and U. L. Gunther, *J Med Chem*, 2008, **51**, 1-3.
97. B. Becattini, C. Culmsee, M. Leone, D. Zhai, X. Zhang, K. J. Crowell, M. F. Rega, S. Landshamer, J. C. Reed, N. Plesnila and M. Pellecchia, *Proc Natl Acad Sci U S A*, 2006, **103**, 12602-12606.
98. L. Poppe, T. S. Harvey, C. Mohr, J. Zondlo, C. M. Tegley, O. Nuanmanee and J. Cheetham, *J Biomol Screen*, 2007, **12**, 301-311.
99. C. Dalvit, M. Flocco, M. Veronesi and B. J. Stockman, *Comb Chem High Throughput Screen*, 2002, **5**, 605-611.
100. C. Dalvit, E. Ardini, M. Flocco, G. P. Fogliatto, N. Mongelli and M. Veronesi, *J Am Chem Soc*, 2003, **125**, 14620-14625.
101. C. Dalvit, A. D. Gossert, J. Coutant and M. Piotto, *Magn Reson Chem*, 2011, **49**, 199-202.
102. N. Kichik, T. Tarrago and E. Giralt, *Chembiochem*, 2010, **11**, 1115-1119.
103. D. A. Pereira and J. A. Williams, *Br J Pharmacol*, 2007, **152**, 53-61.
104. J. R. Miller, V. Thanabal, M. M. Melnick, M. Lall, C. Donovan, R. W. Sarver, D. Y. Lee, J. Ohren and D. Emerson, *Chem Biol Drug Des*, 2010, **75**, 444-454.
105. P. J. Hajduk and J. Greer, *Nat Rev Drug Discov*, 2007, **6**, 211-219.
106. M. J. Gorczynski, J. Grembecka, Y. Zhou, Y. Kong, L. Roudaia, M. G. Douvas, M. Newman, I. Bielnicka, G. Baber, T. Corpora, J. Shi, M. Sridharan, R. Lilien, B. R. Donald, N. A. Speck, M. L. Brown and J. H. Bushweller, *Chem Biol*, 2007, **14**, 1186-1197.
107. D. Gonzalez-Ruiz and H. Gohlke, *J Chem Inf Model*, 2009, **49**, 2260-2271.
108. H. Gohlke, M. Hendlich and G. Klebe, *J Mol Biol*, 2000, **295**, 337-356.
109. K. Ono, H. Ueda, Y. Yoshizawa, D. Akazawa, R. Tanimura, I. Shimada and H. Takahashi, *J Med Chem*, 2010, **53**, 2087-2093.
110. J. Krishnamoorthy, V. C. Yu and Y. K. Mok, *PLoS One*, 2010, **5**, e8943.
111. M. A. McCoy and D. F. Wyss, *J Am Chem Soc*, 2002, **124**, 11758-11763.
112. F. A. Mulder and M. Filatov, *Chem Soc Rev*, 2010, **39**, 578-590.
113. N. E. Englert, C. Richter, J. Wiesner, M. Hintz, H. Jomaa and H. Schwalbe, *Chembiochem*, 2011, **12**, 468-476.
114. V. Tugarinov and L. E. Kay, *J Am Chem Soc*, 2003, **125**, 13868-13878.
115. V. Venditti, N. L. Fawzi and G. M. Clore, *J Biomol NMR*, 2011, **51**, 319-328.
116. G. Pintacuda, M. John, X. C. Su and G. Otting, *Acc Chem Res*, 2007, **40**, 206-212.
117. K. Ogura, T. Shiga, M. Yokochi, S. Yuzawa, T. R. Burke, Jr. and F. Inagaki, *J Biomol NMR*, 2008, **42**, 197-207.
118. S. Han, A. Mistry, J. S. Chang, D. Cunningham, M. Griffor, P. C. Bonnette, H. Wang, B. A. Chrunyk, G. E. Aspnes, D. P. Walker, A. D. Brosius and L. Buckbinder, *J Biol Chem*, 2009, **284**, 13193-13201.
119. Y. Ito and P. Selenko, *Curr Opin Struct Biol*, 2010, **20**, 640-648.

120. D. S. Burz, K. Dutta, D. Cowburn and A. Shekhtman, *Nat Methods*, 2006, **3**, 91-93.
121. J. Xie, R. Thapa, S. Reverdatto, D. S. Burz and A. Shekhtman, *J Med Chem*, 2009, **52**, 3516-3522.
122. Z. Serber, A. T. Keatinge-Clay, R. Ledwidge, A. E. Kelly, S. M. Miller and V. Dotsch, *J Am Chem Soc*, 2001, **123**, 2446-2447.
123. P. Selenko, Z. Serber, B. Gade, J. Ruderman and G. Wagner, *Proc Natl Acad Sci U S A*, 2006, **103**, 11904-11909.
124. T. Takeuchi, M. Kosuge, A. Tadokoro, Y. Sugiura, M. Nishi, M. Kawata, N. Sakai, S. Matile and S. Futaki, *ACS Chem Biol*, 2006, **1**, 299-303.
125. K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki and M. Shirakawa, *Nature*, 2009, **458**, 106-109.
126. M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe and G. Siegal, *Nat Rev Drug Discov*, 2008, **7**, 738-745.
127. K. Kullander and R. Klein, *Nat Rev Mol Cell Biol*, 2002, **3**, 475-486.
128. M. Iiizumi, M. Hosokawa, A. Takehara, S. Chung, T. Nakamura, T. Katagiri, H. Eguchi, H. Ohigashi, O. Ishikawa, Y. Nakamura and H. Nakagawa, *Cancer Sci*, 2006, **97**, 1211-1216.
129. M. Oki, H. Yamamoto, H. Taniguchi, Y. Adachi, K. Imai and Y. Shinomura, *World J Gastroenterol*, 2008, **14**, 5650-5656.
130. K. K. Murai and E. B. Pasquale, *J Cell Sci*, 2003, **116**, 2823-2832.
131. K. L. Binns, P. P. Taylor, F. Sicheri, T. Pawson and S. J. Holland, *Mol Cell Biol*, 2000, **20**, 4791-4805.
132. L. E. Wybenga-Groot, B. Baskin, S. H. Ong, J. Tong, T. Pawson and F. Sicheri, *Cell*, 2001, **106**, 745-757.
133. S. Wiesner, L. E. Wybenga-Groot, N. Warner, H. Lin, T. Pawson, J. D. Forman-Kay and F. Sicheri, *EMBO J*, 2006, **25**, 4686-4696.
134. N. Murali, W. M. Miller, B. K. John, D. A. Avizonis and S. H. Smallcombe, *J Magn Reson*, 2006, **179**, 182-189.
135. P. J. Hajduk, J. R. Huth and S. W. Fesik, *J Med Chem*, 2005, **48**, 2518-2525.
136. S. J. Holland, E. Peles, T. Pawson and J. Schlessinger, *Curr Opin Neurobiol*, 1998, **8**, 117-127.
137. J. P. Labrador, R. Brambilla and R. Klein, *EMBO J*, 1997, **16**, 3889-3897.
138. S. van den Berg, P. A. Lofdahl, T. Hard and H. Berglund, *J Biotechnol*, 2006, **121**, 291-298.
139. W. S. Lewis and S. M. Schuster, *J Biochem Biophys Methods*, 1990, **21**, 129-144.
140. C. J. Asensio and R. C. Garcia, *Anal Biochem*, 2003, **319**, 21-33.
141. J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, G. Superti-Furga, J. Koomen and E. B. Haura, *Nat Chem Biol*, 2010, **6**, 291-299.
142. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 125–132.
143. P. Evans, *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**, 72–82.
144. *Acta Crystallogr. D Biol. Crystallogr.*, 1994, **50**, 760–763.
145. J. Navaza, *Acta Crystallogr. D Biol. Crystallogr.*, 2001, **57**, 1367–1372.

146. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 213–221.
147. P. Emsley and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2126–2132.
148. T. Gronemeyer, C. Chidley, A. Juillerat, C. Heinis and K. Johnsson, *Protein Eng Des Sel*, 2006, **19**, 309-316.
149. G. Siegal and J. G. Hollander, *Curr Top Med Chem*, 2009, **9**, 1736-1745.
150. B. K. Shoichet, *J Med Chem*, 2006, **49**, 7274-7277.
151. A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, 3rd, S. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu and L. M. Shewchuk, *Acta Crystallogr D Biol Crystallogr*, 2007, **63**, 72-79.
152. P. A. Smith, B. C. Tripp, E. A. DiBlasio-Smith, Z. Lu, E. R. LaVallie and J. M. McCoy, *Nucleic Acids Res*, 1998, **26**, 1414-1420.
153. S. L. McGovern and B. K. Shoichet, *J Med Chem*, 2003, **46**, 1478-1483.
154. N. P. Shah, C. Tran, F. Y. Lee, P. Chen, D. Norris and C. L. Sawyers, *Science*, 2004, **305**, 399-401.
155. L. J. Lombardo, F. Y. Lee, P. Chen, D. Norris, J. C. Barrish, K. Behnia, S. Castaneda, L. A. Cornelius, J. Das, A. M. Doweyko, C. Fairchild, J. T. Hunt, I. Inigo, K. Johnston, A. Kamath, D. Kan, H. Klei, P. Marathe, S. Pang, R. Peterson, S. Pitt, G. L. Schieven, R. J. Schmidt, J. Tokarski, M. L. Wen, J. Wityak and R. M. Borzilleri, *J Med Chem*, 2004, **47**, 6658-6661.
156. T. O'Hare, D. K. Walters, E. P. Stoffregen, T. Jia, P. W. Manley, J. Mestan, S. W. Cowan-Jacob, F. Y. Lee, M. C. Heinrich, M. W. Deininger and B. J. Druker, *Cancer Res*, 2005, **65**, 4500-4505.
157. W. Kabsch, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 125-132.
158. P. Evans, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 72-82.
159. *Acta Crystallogr D Biol Crystallogr*, 1994, **50**, 760-763.
160. J. Navaza, *Acta Crystallogr D Biol Crystallogr*, 2001, **57**, 1367-1372.
161. G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis, *Nat Protoc*, 2008, **3**, 1171-1179.
162. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr D Biol Crystallogr*, 1997, **53**, 240-255.
163. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr D Biol Crystallogr*, 2010, **66**, 213-221.
164. P. Emsley and K. Cowtan, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 2126-2132.
165. A. P. Kornev, S. S. Taylor and L. F. Ten Eyck, *Proc Natl Acad Sci U S A*, 2008, **105**, 14377-14382.

166. S. R. Hubbard, M. Mohammadi and J. Schlessinger, *J Biol Chem*, 1998, **273**, 11987-11990.
167. M. Huse and J. Kuriyan, *Cell*, 2002, **109**, 275-282.
168. S. W. Cowan-Jacob, H. Mobitz and D. Fabbro, *Curr Opin Cell Biol*, 2009, **21**, 280-287.
169. (The PyMOL Molecular Graphics System; DeLano Scientific, 2002).
170. J. Nowakowski, C. N. Cronin, D. E. McRee, M. W. Knuth, C. G. Nelson, N. P. Pavletich, J. Rogers, B. C. Sang, D. N. Scheibe, R. V. Swanson and D. A. Thompson, *Structure*, 2002, **10**, 1659-1667.
171. L. E. Kay, D. R. Muhandiram, N. A. Farrow, Y. Aubin and J. D. Forman-Kay, *Biochemistry*, 1996, **35**, 361-368.
172. Y. Liu, K. Shah, F. Yang, L. Witucki and K. M. Shokat, *Bioorg Med Chem*, 1998, **6**, 1219-1226.
173. H. Daub, K. Specht and A. Ullrich, *Nat Rev Drug Discov*, 2004, **3**, 1001-1010.
174. M. Getlik, C. Grutter, J. R. Simard, S. Kluter, M. Rabiller, H. B. Rode, A. Robubi and D. Rauh, *J Med Chem*, 2009, **52**, 3915-3926.
175. N. K. Williams, I. S. Lucet, S. P. Klinken, E. Ingley and J. Rossjohn, *J Biol Chem*, 2009, **284**, 284-291.
176. J. S. Tokarski, J. A. Newitt, C. Y. Chang, J. D. Cheng, M. Wittekind, S. E. Kiefer, K. Kish, F. Y. Lee, R. Borzillerri, L. J. Lombardo, D. Xie, Y. Zhang and H. E. Klei, *Cancer Res*, 2006, **66**, 5790-5797.
177. D. J. Marcotte, Y. T. Liu, R. M. Arduini, C. A. Hession, K. Miatkowski, C. P. Wildes, P. F. Cullen, V. Hong, B. T. Hopkins, E. Mertsching, T. J. Jenkins, M. J. Romanowski, D. P. Baker and L. F. Silvian, *Protein Sci*, 2010, **19**, 429-439.
178. J.-P. Himanen, M. J. Chumley, M. Lackmann, C. Li, W. A. Barton, P. D. Jeffrey, C. Vearing, D. Geleick, D. A. Feldheim, A. W. Boyd, M. Henkemeyer and D. B. Nikolov, *Nat. Neurosci.*, 2004, **7**, 501–509.
179. K. Kullander and R. Klein, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 475–486.
180. R. van Doorn, R. Dijkman, M. H. Vermeer, J. J. Out-Luiting, E. M. van der Raaij-Helmer, R. Willemze and C. P. Tensen, *Cancer Res.*, 2004, **64**, 5578–5586.
181. M. Iiizumi, M. Hosokawa, A. Takehara, S. Chung, T. Nakamura, T. Katagiri, H. Eguchi, H. Ohigashi, O. Ishikawa, Y. Nakamura and H. Nakagawa, *Cancer Sci.*, 2006, **97**, 1211–1216.
182. M. Oki, H. Yamamoto, H. Taniguchi, Y. Adachi, K. Imai and Y. Shinomura, *World J. Gastroenterol.*, 2008, **14**, 5650–5656.
183. S. Ashida, H. Nakagawa, T. Katagiri, M. Furihata, M. Iiizumi, Y. Anazawa, T. Tsunoda, R. Takata, K. Kasahara, T. Miki, T. Fujioka, T. Shuin and Y. Nakamura, *Cancer Res.*, 2004, **64**, 5963–5972.
184. Y. Goldshmit, M. P. Galea, G. Wise, P. F. Bartlett and A. M. Turnley, *J. Neurosci.*, 2004, **24**, 10064–10073.
185. J. Fabes, P. Anderson, C. Brennan and S. Bolsover, *Eur. J. Neurosci.*, 2007, **26**, 2496–2505.
186. P. Bamborough, D. Drewry, G. Harper, G. K. Smith and K. Schneider, *J. Med. Chem.*, 2008, **51**, 7898–7914.
187. M. W. Karaman, S. Herrgard, D. K. Treiber, P. Gallant, C. E. Atteridge, B. T. Campbell, K. W. Chan, P. Ciceri, M. I. Davis, P. T. Edeen, R. Faraoni, M. Floyd,

J. P. Hunt, D. J. Lockhart, Z. V. Milanov, M. J. Morrison, G. Pallares, H. K. Patel, S. Pritchard, L. M. Wodicka and P. P. Zarrinkar, *Nat. Biotechnol.*, 2008, **26**, 127–132.
188. R. Noberini, M. Koolpe, S. Peddibhotla, R. Dahl, Y. Su, N. D. P. Cosford, G. P. Roth and E. B. Pasquale, *J. Biol. Chem.*, 2008, **283**, 29461–29472.
189. J. J. Liao, *J. Med. Chem.*, 2007, **50**, 409–424.
190. C. Farenc, P. H. Celie, C. P. Tensen, I. J. de Esch and G. Siegal, *FEBS Lett*, 2011.
191. S. Wiesner, L. E. Wybenga-Groot, N. Warner, H. Lin, T. Pawson, J. D. Forman-Kay and F. Sicheri, *EMBO J.*, 2006, **25**, 4686–4696.
192. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, **298**, 1912–1934.
193. T. L. Davis, J. R. Walker, P. Loppnau, C. Butler-Cole, A. Allali-Hassani and S. Dhe-Paganon, *Structure*, 2008, **16**, 873–884.
194. J. S. Tokarski, J. A. Newitt, C. Y. J. Chang, J. D. Cheng, M. Wittekind, S. E. Kiefer, K. Kish, F. Y. F. Lee, R. Borzillerri, L. J. Lombardo, D. Xie, Y. Zhang and H. E. Klei, *Cancer Res.*, 2006, **66**, 5790–5797.
195. M. J. McGregor, *J. Chem. Inf. Model.*, 2007, **47**, 2374–2382.
196. F. Zuccotto, E. Ardini, E. Casale and M. Angiolini, *J. Med. Chem.*, 2009, **53**, 2681–2694.
197. A. R. Leach, B. K. Shoichet and C. E. Peishoff, *J. Med. Chem.*, 2006, **49**, 5851–5855.
198. S. W. Cowan-Jacob, H. Mobitz and D. Fabbro, *Curr. Opin. Cell Biol.*, 2009, **21**, 280–287.
199. O. P. van Linden, C. Farenc, W. H. Zoutman, L. Hameetman, M. Wijtmans, R. Leurs, C. P. Tensen, G. Siegal and I. J. de Esch, *Eur J Med Chem*, 2012, **47**, 493-500.
200. H. A. Rajapakse, K. P. Moore, P. G. Nantermet, J. M. Sanders, S. Parmentier-Batteur and R. J. Mark, *US20100113415.*, 2010.
201. S. Parmentier-Batteur, E. N. Finger, R. Krishnan, H. A. Rajapakse, J. M. Sanders, G. Kandpal, H. Zhu, K. P. Moore, C. P. Regan, S. Sharma, J. Fred Hess, T. M. Williams, I. J. Reynolds, J. P. Vacca, R. J. Mark and P. G. Nantermet, *J. Neurochem.*, 2011, Accepted Article.
202. J. Fabes, P. Anderson, C. Brennan and S. Bolsover, *Eur J Neurosci*, 2007, **26**, 2496-2505.
203. M. Congreve, G. Chessari, D. Tisi and A. J. Woodhead, *J Med Chem*, 2008, **51**, 3661-3680.
204. M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle and H. Jhoti, *J Med Chem*, 2005, **48**, 403-413.
205. Z. K. Sweeney, S. Acharya, A. Briggs, J. P. Dunn, T. R. Elworthy, J. Fretland, A. M. Giannetti, G. Heilek, Y. Li, A. C. Kaiser, M. Martin, Y. D. Saito, M. Smith, J. M. Suh, S. Swallow, J. Wu, J. Q. Hang, A. S. Zhou and K. Klumpp, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 4348-4351.
206. D. G. Myszka, *Methods Enzymol*, 2000, **323**, 325-340.
207. S. Perspicace, D. Banner, J. Benz, F. Muller, D. Schlatter and W. Huber, *J Biomol Screen*, 2009, **14**, 337-349.
208. T. D. K. Goddard, D.G., *SPARKY 3, University of California, San Francisco*.

209. A. M. Giannetti, B. D. Koch and M. F. Browner, *J Med Chem*, 2008, **51**, 574-580.
210. A. M. Giannetti, *Methods Enzymol*, 2011, **493**, 169-218.
211. A. P. Van Der Merwe, *Oxford, London, p 137*, 2001.
212. P. Pattnaik, *Appl Biochem Biotechnol*, 2005, **126**, 79-92.
213. H. Nordin, M. Jungnelius, R. Karlsson and O. P. Karlsson, *Anal Biochem*, 2005, **340**, 359-368.
214. S. J. Pollack, K. S. Beyer, C. Lock, I. Muller, D. Sheppard, M. Lipkin, D. Hardick, P. Blurton, P. M. Leonard, P. A. Hubbard, D. Todd, C. M. Richardson, T. Ahrens, M. Baader, D. O. Hafenbradl, K. Hilyard and R. W. Burli, *J Comput Aided Mol Des*, 2011, **25**, 677-687.
215. A. Medek, P. J. Hajduk, J. Mack and S. W. Fesik, *J Am Chem Soc*, 2000, **122**, 1241-1242.
216. C. Dalvit, *Drug Discov Today*, 2009, **14**, 1051-1057.
217. B. Heras and J. L. Martin, *Acta Crystallogr D Biol Crystallogr*, 2005, **61**, 1173-1180.

Summary

In multicellular organisms, communication between individual cells is essential for the regulation and coordination of complex cellular processes such as gene expression, metabolic pathways, cell growth and differentiation. Receptors tyrosine kinases or RTKs are cell surface receptors that regulate such cellular processes, but also have a critical role in the development and progression of many types of cancer. The overexpression of EphA4, a member of the RTK family, has been observed in a variety of malignant carcinomas including gastric cancer and cutaneous lymphomas among others. Therefore, the aim of the research project that is associated with this thesis was to develop high affinity inhibitors of the tyrosine kinase EphA4.

Fragment-based drug discovery (FBDD) has been particularly successful in developing kinase inhibitors. Chapter 1 presents the drug discovery process and introduces the notion of FBDD. Historically, most pharmaceutical companies have used combinatorial chemistry and high throughput screening (HTS) to identify active molecules. In the past decades, FBDD has emerged as an alternative method used for finding lead compounds. This approach is based on identifying small molecules, or fragments (< 300 Da), which may bind only weakly to the biological target, and then growing them or combining them to produce a lead with a higher affinity. To detect weak binding of fragments to a protein, biophysical methods such as Nuclear Magnetic Resonance (NMR) are required. Chapter 2 presents an overview of biomolecular NMR in drug discovery. From ligand discovery to in cell applications, NMR is used at each

stage of the drug discovery process. Due to the sensitivity of the technique, NMR is the most frequently used method of fragment screening.

In this project, ligand discovery was based on two complementary approaches, a computational screen and an NMR based screen using Target Immobilized NMR Screening (TINS). The TINS technology involves immobilization on resin of a target protein and a reference protein with minimal small molecule binding properties. Identification of fragments which bind specifically to the target can be done by simple comparison of the 1D ^{1}H signal intensities of the fragments in the presence of the reference or target. In Chapter 3, the TINS approach was applied to EphA4. A portion of the receptor, the kinase domain which contains the catalytic function, was chosen as the target protein. Several immobilization strategies were investigated in order to yield a functional protein on resin and the target was subsequently screened. The hits obtained were biochemically characterized and selected compounds were submitted to crystallographic structure determination. The biochemical data yielded a surprisingly low validation rate of the compounds, and crystallography failed to give structures of fragments in complex with EphA4. Thus an alternative immobilization was investigated and an additional TINS screen was performed. Furthermore, analysis of putative fragment binding using orthogonal biophysical methods including Surface Plasmon Resonance (SPR) and protein observed NMR was employed as described in Chapter 6.

The crystal structure of the EphA4 kinase domain was solved in Chapter 4. In addition, the structure of the kinase domain in complex with dasatinib, a well-known kinase inhibitor, was also elucidated. The two structures were examined and compared with the crystals structures of other Eph kinase domains. This analysis revealed a

hydrophobic back pocket in the ATP binding site of EphA4 kinase domain which was unknown before. This newly discovered pocket, when followed up, could lead to inhibitor specificity during drug design.

In parallel to the NMR based screen, a computational approach towards EphA4 was performed by the Ph.D. candidate Oscar van Linden in the group of Dr. de Esch in Amsterdam (Chapter 5). This *in silico* approach yielded an interesting and potent inhibitor of the kinase domain. This compound was biophysically characterized using SPR in Chapter 6. Furthermore, soaking experiments into crystals of the native EphA4 were undertaken and the binding mode was elucidated (Chapter 5).

The same soaking procedure was employed with the compounds originating from the TINS screen, however none of the fragments yielded a structure (Chapter 3). As the crystallographic conditions are far from physiological pH, [^{15}N,^{1}H]-HSQC NMR experiments were conducted to test ligand binding at the pH used for crystallography in order to investigate whether this difference could be the cause for the lack of crystal structures (Chapter 6). Results show that, for the compounds assessed, the change in pH does not preclude compound binding.

Biophysical characterization of the compounds previously identified in Chapter 3 was conducted using SPR technology (Chapter 6). Several immobilization strategies were investigated and a good correlation was observed between the results from the TINS screen and the SPR analysis. In addition, two fragments were assessed for competition binding using ATP, and were shown to bind in the ATP binding site. Although the binding mode could not be determined via X- ray crystallography, the similarity of

structures between one of these fragment and the potent inhibitor discovered via *in silico* suggested a similar binding mode.

Finally, Chapter 7 describes the conclusions and perspectives of this thesis. This research project led to the identification of several compounds that were biochemically and biophysically characterized. The *in silico* approach discovered a potent inhibitor of EphA4 for which the binding mode was elucidated via X-ray crystallography. Moreover, the TINS approach identified two compounds that may constitute starting points for the generation of more potent EphA4 inhibitors. Further improvement of these compounds could lead to the development of a new drug for the treatment of cancer and neuronal injuries by inhibition of the EphA4 receptor tyrosine kinase.

Samenvatting

Communicatie tussen individuele cellen in meercellige organismen is essentieel voor de regulatie en coördinatie van complexe cellulaire processen zoals gen expressie, metabolisme, cel groei en differentiatie. Receptor Tyrosine Kinases of RTK's zijn receptoren die aan de cel oppervlakte dit soort cel processen reguleren, maar ze spelen ook een kritieke rol in de ontwikkeling en vooruitgang van diverse soorten kanker. De over-expressie van EphA4, lid van de RTK familie, is waargenomen in diverse kwaadaardige carcinomen zoals onder andere darmkanker en huidlymfomen. Daarom is het doel van het onderzoek gerelateerd aan dit proefschrift het ontwikkelen van hoge affiniteit remmers voor tyrosine kinase EphA4.

Fragment –based drug discovery (FBDD) is een succesvolle techniek gebleken voor de ontwikkeling van kinase remmers. Hoofdstuk 1 beschrijft het proces van medicijn ontwikkeling en introduceert het begrip FBDD. In het verleden hebben de meeste farmaceutische bedrijven gebruik gemaakt van combinatorische chemie en experimenten met een hoge doorvoer oftewel *high throughput screening* (HTS) om actieve moleculen te identificeren. De laatste jaren is FBDD naar voren gekomen als alternatieve methode om interessante moleculen te vinden. Deze methode is gebaseerd op de identificatie van kleine moleculen of fragmenten (<300 Da), die misschien maar zwak aan het biologische doelwit binden, om ze vervolgens uit te breiden of the combineren tot een molecuul met een hogere affiniteit. Om de zwakke binding van fragmenten aan een eiwit te detecteren zijn biofysische methoden zoals kernspinresonantie (*Nuclear Magnetic Resonance*, NMR) noodzakelijk. Hoofdstuk 2 geeft een overzicht van

biomoleculaire NMR in geneesmiddelenontwikkeling. Van de ontdekking van liganden tot “in cel” applicaties, NMR wordt gebruikt binnen elke fase van het proces van geneesmiddelenontwikkeling. Vanwege de gevoeligheid van de techniek, is NMR de meest gebruikte methode voor het screenen van fragmenten.

In dit project zijn twee complementaire methodes gebruikt voor het vinden van liganden: een computersimulatie en een NMR gebaseerde methode, *Target Immobilized NMR screening* (TINS). Bij TINS technologie worden het eiwit van interesse en een controle eiwit, met de eigenschap minimale binding van kleine moleculen te vertonen, geïmmobiliseerd op sepharose bolletjes. Identificatie van fragmenten die specifiek aan het eiwit van interesse binden vindt plaats door simpele vergelijking van de 1D ^{1}H signaal intensiteiten van de fragmenten in aanwezigheid van eiwit van interesse of controle eiwit. In hoofdstuk 3 wordt beschreven hoe de TINS technologie toegepast wordt op EphA4. Een deel van de receptor, het kinase domein met katalytische functie, is gekozen als eiwit van interesse. Verschillende immobilisatie technieken zijn onderzocht teneinde functioneel eiwit aan sepharose bolletjes te verkrijgen. De verkregen treffers zijn biochemisch gekarakteriseerde en geselecteerde organische verbindingen, die zijn gebruikt voor structuur bepalingen middels kristallografie. De biochemische karakterisatie leverde een verrassend lage validatie ratio op van deze organische verbindingen en kristallisatie heeft niet geleidt tot een structuur van fragmenten in complex met EphA4. Om deze reden is een alternatieve wijze van immobilisatie onderzocht en een tweede TINS experiment uitgevoerd. Daarnaast is vermeende fragment binding geanalyseerd door gebruik te maken van rechtstreekse biofysische

methoden zoals *Surface Plasmon Resonance* (SPR) en eiwit observatie in NMR zoals beschreven in hoofdstuk 6.

Het oplossen van kristalstructuur van het EphA4 kinase domein is beschreven in hoofdstuk 4. Daarnaast is de structuur opgehelderd van het kinase domein in complex met dasatinib, een bekende kinase-remmer. De twee structuren zijn onderzocht en vergeleken met de kristalstructuren van andere Eph kinase domeinen. Deze analyse onthulde een hydrofobe holte in de ATP bindingsplaats van het EphA4 kinase domein die nog niet eerder bekend was. Deze nieuw ontdekte holte zou bij nader onderzoek kunnen leiden tot specificiteit van een remmer tijdens het proces van geneesmiddelenontwikkeling.

Naast het NMR onderzoek, is er ook een gecomputeriseerde aanpak gebruikt richting EphA4 door promovendus Oscar van Linden uit de groep van Dr. de Esch in Amsterdam (hoofdstuk 5). Deze *in silico* aanpak leverde een interessante en potentiële remmer op van het kinase domein. Deze organische verbinding is biofysisch gekarakteriseerd door middel van SPR zoals vermeld in hoofdstuk 6. Daarnaast zijn experimenten uitgevoerd waarbij de kristallen van het natieve EphA4 werden geïncubeerd met de organische verbindingen en is de bindingswijze verhelderd.

Dezelfde methode van incubatie is gebruikt met de organische verbindingen die naar voren kwamen tijdens het TINS experiment; helaas leverde geen van deze fragmenten een bruikbare structuur (hoofdstuk 3). De omstandigheden tijdens kristallografie verschillen zeer van de fysiologische pH. Hierom zijn [^{15}N,^{1}H]-HSQC NMR experimenten uitgevoerd om te onderzoeken of dit verschil de oorzaak kan zijn

voor het gebrek aan kristalstructuren (hoofdstuk 6). Resultaten van de beoordeelde organische verbindingen laten zien dat een verandering in pH binding niet uit sluit.

Biofysische karakterisatie van de verbindingen die eerder werden geïdentificeerd zoals beschreven in hoofdstuk 3 is uitgevoerd door gebruik te maken van SPR technologie (hoofdstuk 6). Meerdere strategieën voor immobillisatie zijn onderzocht en er werd een goede correlatie gevonden tussen de resultaten uit het TINS experiment en de SPR analyse. Daarnaast zijn twee organische verbindingen beoordeeld op concurrerende binding waarbij gebruik werd gemaakt van ATP en waarvan is aangetoond dat deze de ATP bindingsplek binden. Hoewel de wijze van binding niet kon worden bepaald via kristallografie, doet de overeenkomst in structuur tussen één van deze organische verbindingen en de potentiële remmer eenzelfde wijze van binding vermoeden.

Ten slotte beschrijft hoofdstuk 7 de conclusies en vooruitzichten van dit proefschrift. Dit onderzoeksproject heeft geleidt tot de identificatie van meerdere organische verbindingen die biochemisch en biofysisch zijn gekarakteriseerd. De *in silico* aanpak leverde een redelijk potentiële remmer van EphA4 op waarvan de bindingswijze via kristallografie is verhelderd. Bovendien heeft de TINS aanpak twee organische verbindingen geïdentificeerd die een startpunt kunnen vormen voor ontwikkeling van een meer potente remmer van EphA4. Verdere verbetering van deze organische verbindingen kan leiden tot de ontwikkeling van een nieuw geneesmiddel voor de behandeling van kanker en neuraal letsel door remming van de EphA4 receptor tyrosine kinase.

Résumé

Au sein des organismes multicellulaires, la communication intercellulaire s'avère essentielle pour la régulation et la coordination de processus cellulaires complexes, tels que l'expression des gènes, la voie métabolique, ou encore la croissance et la différenciation des cellules. Les récepteurs à activité tyrosine kinase ou «RTK» sont des récepteurs situés à la surface des cellules, qui non seulement régulent de tels processus mais jouent également un rôle fondamental dans le développement et la progression de nombreux types de cancer ; on a ainsi observé une surexpression du récepteur EphA4 — de la famille des RTK — dans différents carcinomes malins dont, entre autres, des cancers gastriques et des lymphomes cutanés. L'objectif du projet de recherche associé à cette thèse était donc de développer des inhibiteurs haute affinité du récepteur à activité tyrosine kinase EphA4.

La conception de médicament basé sur des fragments de molécules ou «FBBD» *(fragment-based drug discovery)* s'est révélée particulièrement fructueuse en ce qui concerne le développement d'inhibiteurs de kinase. Le chapitre 1 présente le processus de conception de médicament et introduit la notion de FBBD. De tous temps, la plupart des entreprises pharmaceutiques ont eu recours à la chimie combinatoire et au criblage à haut débit (HTS : High Throughput Screening) pour identifier des molécules actives. Au cours des dernières décennies, la FBBD s'est révélé être une méthode alternative pour la recherche de composés phares pouvant être a l'origine de médicaments. Cette approche s'appuie sur l'identification de petites molécules, ou fragments (<300 Da), qui se lient faiblement à la cible biologique, pour ensuite les faire croître ou les combiner pour produire une molécule ayant une affinité plus élevée. Des méthodes biophysiques, telles que la résonance magnétique nucléaire (RMN), sont nécessaires pour détecter de faibles affinités de liaison. Le chapitre 2 traite des tenants et aboutissants de la

RMN biomoléculaire dans le domaine de la conception de médicament. De la découverte de ligand aux applications intracellulaires, la RMN intervient à chaque étape du processus de conception de médicament et, en raison de sa sensibilité, est la méthode la plus usitée pour le criblage de fragment.

Dans ce projet, deux méthodes complémentaires de modélisation ont sous-tendu la découverte de ligand : une première assistée par ordinateur et une seconde basée sur la RMN au moyen d'un criblage RMN sur une protéine cible immobilisée ou «TINS» *(Target Immobilized NMR Screening)*. La technologie TINS implique l'immobilisation sur résine de la protéine cible et d'une protéine témoin ayant des propriétés de liaison intermoléculaire minimales. L'identification de fragments se fixant spécifiquement à la cible se fait ainsi par la simple comparaison des intensités du signal 1D ^{1}H des fragments en présence de la protéine témoin, puis de la protéine cible. Le chapitre 3 est consacré à l'application de la méthode TINS au récepteur EphA4. Une partie de celui-ci, à savoir le domaine kinase responsable de la fonction catalytique, se vit conférer le rôle de protéine cible. Plusieurs stratégies d'immobilisation ont été tentées afin d'obtenir un échantillon de protéine sur résine utilisable, la cible étant par la suite modélisée. Les résultats obtenus ont été biochimiquement caractérisés, et la structure cristallographique des composés obtenus déterminée. Les données biochimiques ont livré un taux de validation étonnamment faible, et la cristallographie n'a pas permis de déterminer des structures de fragments susceptibles de former un complexe avec le récepteur EphA4. Une méthode d'immobilisation alternative a alors été recherchée, et une modélisation TINS supplémentaire réalisée. Il a été en outre mené une analyse des liaisons des fragments putatifs, décrite au chapitre 6, au moyen de méthodes biophysiques orthogonales, dont la résonance plasmon de surface (SPR) et la RMN aux fins d'observation de protéine.

La structure cristalline du domaine kinase du récepteur EphA4 a été déterminée au chapitre 4. Il en a été de même de la structure du domaine kinase formant un complexe avec le

dasatinib, qui est un inhibiteur de kinase bien connu. Les deux structures ont fait l'objet d'un examen approfondi et d'une comparaison avec les structures cristallines des domaines kinases d'autres récepteurs Eph. Cette étude a révélé une partie hydrophobe — inconnue jusqu'alors — dans le site de liaison ATP du domaine kinase. L'étude de cette face nouvellement découverte a montré qu'il pourrait ouvrir la voie à une spécificité inhibitrice au cours de la conception de médicament.

En parallèle de la modélisation basée sur la RMN, Oscar van Linden, doctorant faisant partie de l'équipe du Dr. de Esch à Amsterdam, a réalisé une modélisation informatique du récepteur EphA4 (chapitre 5). Cette méthode *in silico* a produit un puissant et intéressant inhibiteur du domaine kinase. Le chapitre 6 explique comment les caractéristiques biophysiques de ce composé ont été déterminées au moyen de la résonance plasmon de surface (SPR). De plus, des expériences de trempage de cristaux natifs dans une solution contenant cet inhibiteur ont permis de résoudre le mode de liaison de l'inhibiteur à la protéine. (chapitre 5).

La même procédure a été utilisée sur les composés issus du criblage TINS, mais sans qu'aucun des fragments ne livre de structure (chapitre 3). Les conditions cristallographiques étant éloignées du pH physiologique, des expériences s'appuyant sur la RMN [^{15}N,^{1}H]-HSQC ont été menées afin de déterminer si cette différence pouvait être la cause de l'absence de structures cristallines (chapitre 6). Les résultats montrent que, pour ce qui est des composés examinés, la différence de pH n'empêche pas la liaison.

Les caractéristiques biophysiques des composés précédemment identifiés au chapitre 3 ont été définis en ayant recours à la technologie RPS (chapitre 6). Plusieurs stratégies d'immobilisation ont été tentées, et une bonne corrélation a pu être observée entre les résultats fournis, d'une part, par la modélisation TINS et, d'autre part, par l'analyse RPS. En outre, deux fragments ont été soumis à un test de liaison en utilisant de l'ATP et se sont effectivement

attachés sur le site de liaison ATP. Bien que le mode de liaison n'eût pu être déterminée par une cristallographie aux rayons X, la similarité des structures entre un de ces fragments et le puissant inhibiteur découvert *in silico* semblait indiquer un mode de liaison analogue.

Enfin, le chapitre 7 présente les conclusions et perspectives de cette thèse. Ce projet de recherche a abouti à l'identification de plusieurs composés dont les caractéristiques biochimiques et biophysiques ont été déterminées. La méthode *in silico* a permis la découverte d'un inhibiteur de l'EphA4 relativement puissant dont le mode de liaison a été décrypté en utilisant la cristallographie aux rayons X. De plus, la méthode TINS a permis d'identifier deux composés qui peuvent être considérés comme les prémices d'une génération d'inhibiteurs EPhA4 plus puissants. Une version plus aboutie de ces composés pourrait mener au développement d'un nouveau médicament pour le traitement du cancer et des dommages neurologiques par inhibition du récepteur à activité tyrosine kinase EphA4.

Curriculum Vitae

Carine Farenc was born on 10th of January 1981 in Albi, France. She grew up in Camarès and pursued her secondary education in Saint-Affrique. After obtaining her Baccalaureate degree, she moved to Toulouse where she obtained her Bachelor in Biochemistry and Molecular Biology. Then, she pursued her studies in biochemistry and structural biology for one year in the program Master of Biochemistry and Biotechnologies and a further one year obtaining a Professional Master in Structural Biochemistry and Proteomics. The last seven months of the master's course were spent at Astrazeneca in Alderley Park UK, where she pursued research in structural biology and especially in drug discovery. In 2007 she started to work as a PhD student under the supervision of Dr. Gregg Siegal in the ProtChem group of Pr. Dr. Marcellus Ubbink at the Leiden Institute of Chemistry. The research focused mainly on developing high affinity inhibitors of the tyrosine kinase EphA4 using fragment based drug discovery. To reach this aim, several techniques such as NMR, SPR and other biochemical methods were employed.

As of March 2012, Carine is employed as a post-doctoral researcher at the AFMB in Marseille in the group of Dr. Christian Cambillau.

List of publications

C. Farenc and G. Siegal (2012) Recent Advances in Biomolecular NMR for Drug Discovery.
Chapter from "Recent Advances in Biomolecular NMR" in the RSC Biomolecular Sciences Series. 2012

C. Farenc, P. H. Celie, C.P. Tesssen, I.J. de Esch and G. Siegal (2011) Crystal structure of the EphA4 protein tyrosine kinase domain in the apo- and dasatinib-bound state. *FEBS Letters* 585, 3593-3599.

O. P. van Linden, C. Farenc, W. H. Zoutman, L. Hameetman, M. Wijtmans, R. Leurs, C. P. Tensen, G. Siegal and I. J. de Esch (2012) Fragment based lead discovery of small molecule inhibitors for the EPHA4 receptor tyrosine kinase. *Eur J Med Chem* 47, 493-500.